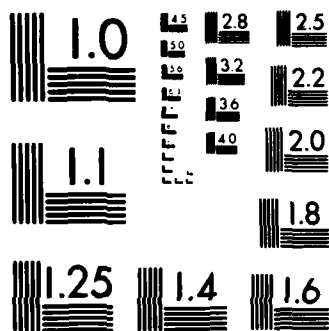END

FILMED

DTIC

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

BOUNDARY CROSSING PROBABILITIES
AND STATISTICAL APPLICATIONS
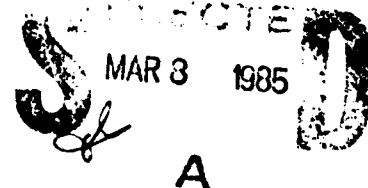
by

David Siegmund
Stanford University

TECHNICAL REPORT NO. 32
JANUARY 1985

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

MAR 3 1985

A

85

BOUNDARY CROSSING PROBABILITIES
AND STATISTICAL APPLICATIONS

by

David Siegmund
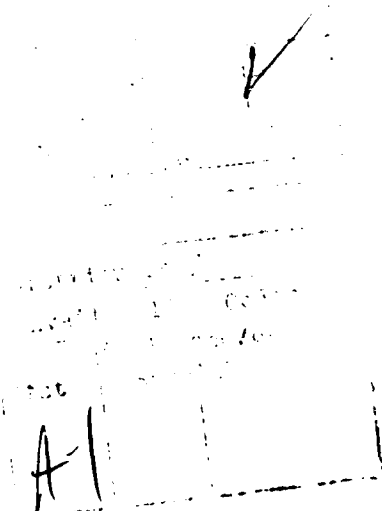Stanford University

TECHNICAL REPORT NO. 32
JANUARY 1985

PREPARED UNDER CONTRACT
N00014-77-C-0306 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Boundary Crossing Probabilities and Statistical Applications

David Siegmund
Stanford Univeristy

Abstract

This paper surveys recent results involving boundary crossing probabilities and related statistical applications. The first part is concerned with problems of sequential analysis, especially repeated significance tests and their application to sequential clinical trials involving survival data. The second part develops the probability theory motivated by the problems of Part 1. A method for computing first passage distributions of Brownian motion to linear boundaries is introduced and then modified to handle problems in discrete time and those involving nonlinear boundaries. The third part is concerned with fixed sample statistical problems, especially change-point problems, which involve boundary crossing probabilities. Examples are given of problems for which the methods of Part 2 appear adequate and of problems which require new methods.

Boundary Crossing Probabilities and Statistical Applications

David Siegmund
Stanford University

0. Introduction

Let $X(t)$, $t = 1, 2, \cdots$ or $0 \leq t < \infty$, be a stochastic process and let $c(t)$ be constants. The general subject of this paper is approximate computation of boundary crossing probabilities of the form

$$(0.1) \qquad P\{X(t) \geq c(t) \quad \text{for some} \quad m_0 \leq t \leq m\}$$

or

$$(0.2) \qquad P\{X(t) \geq c(t) \quad \text{for some} \quad m_0 \leq t \leq m \mid X(m) = \xi\}$$

and statistical applications of the resulting approximations.

The grandfather of all such problems in statistics is to determine the distribution of the one sample Kolmogorov-Smirnov statistic, which is of the form (0.1) with $X(t)$ the difference between the empirical and true distribution function of a random sample and $c(t)$ identically constant. The limiting distribution of this statistic is of the form (0.2) with $X(t)$ a Brownian motion process, $c(t)$ identically constant, and $\xi = 0$. The principal contemporary motivation for studying such problems comes from sequential analysis, which is the context in which many of the results discussed below first arose.

The paper is divided into three parts. The first is concerned with a class of problems in sequential analysis which lead naturally to problems of the form (0.1) and (0.2). The discussion in Part 1 is restricted to statistical issues. We shall in effect assume that we can compute without difficulty various boundary crossing probabilities that arise. However, these problems motivate the second part of the paper, which is concerned with the mathematical problem of approximation of boundary crossing probabilities. The third part discusses a number of non-sequential statistical problems which also lead to boundary crossing probabilities, some of which are essentially already solved in Part 2, and some of which

1

require the development of new methods. Particular attention is given to so-called "change point" problems. The results here are in some respects less complete and outline a program for future research.

Because the subject of boundary crossing probabilities is quite technical, to convey the main ideas the following discussion is frequently heuristic and restricted to special cases. References are given to mathematically rigorous treatments. I have written a monograph on sequential analysis (Siegmund, 1985), which describes in substantially more detail most of the results of Parts 1 and 2.

1. Sequential Analysis

The primary impetus for the development of sequential analysis during the 1940's was a desire for more efficient methods of sampling inspection. Recent developments have been motivated at least in part by ethical considerations in the design of clinical trials.

1.1 Repeated Significance Tests for Normal Data

We shall consider in detail the following very simple model of a clinical trial. In order to compare two treatments, A and B, patients arrive sequentially and are paired, with one patient in each pair receiving treatment A and the other treatment B. Let a_i (b_i) denote the (immediate) response of the patient in the ith pair who receives treatment A (B), and let $z_i = a_i - b_i$. Assume that $z_1, z_2, \cdots$ are independent and normally distributed random variables with mean μ and known variance, which without loss of generality can be assumed equal to 1. Our primary goal is to test the hypothesis of no treatment effect, $H_0 : \mu = 0$, against the alternative $H_1 : \mu \neq 0$. Of course the standard fixed sample size test (at level .05) based on a sample of size n is to compute $S_n = z_1 + \cdots + z_n$ and reject H_0 if $|S_n| \geq b'n^{1/2}$ ($b' = 1.96$).

If μ is considerably different from 0, indicating that one of the two treatments is considerably superior to the other, it is desirable to ascertain this fact with a minimum amount of experimentation, so that all future patients can receive the (apparently) superior treatment. On the other hand if μ is about equal to 0, there is no ethical mandate (although

2

there may be a financial one) to stop sampling as soon as possible. A sequential test designed to stop sampling as soon as it is apparent that H_1 is true while behaving like a fixed sample test if H_0 appears to be true is the so-called repeated significance test, defined as follows (cf. Armitage, 1975).

Given m_0, m, and $b > 0$, define the stopping rule

$$(1.1) \qquad T = \inf\{n : n \geq m_0, \ |S_n| \geq bn^{1/2}\}.$$

Stop sampling at $\min(T, m)$ and reject H_0 if and only if $T \leq m$. The power of this test is

$$(1.2) \qquad P_\mu\{T \leq m\} = P_\mu\left(\bigcup_{n=m_0}^{m} \{|S_n| \geq bn^{1/2}\}\right).$$

Its expected sample size is

$$(1.3) \qquad E_\mu[\min(T, m)],$$

which we anticipate will be small when $|\mu| \gg 0$ and about equal to m when $\mu \cong 0$.

Remark: Note that the stopping rule (1.1) can be written

$$T = \inf\{n : n \geq m_0, \quad S_n^2/2n \geq b^2/2\},$$

or

$$(1.4) \qquad T = \inf\{n : n \geq m_0, \ \Lambda_n \geq a\},$$

where Λ_n is the log likelihood ratio statistic for testing $\mu = 0$ against $\mu \neq 0$, and $a = b^2/2$. This observation is very helpful in adapting the results developed here to different situations. For example, if we drop the hypothesis that the variance of the z's is known, (1.4) becomes

$$T = \inf\{n : n \geq m_0, \quad (n/2)\log(1 + \bar{z}_n^2/v_n^2) \geq a\},$$

where $\bar{z}_n = n^{-1}\sum_1^n z_i$ and $v_n^2 = n^{-1}\sum_1^n(z_i - \bar{z}_n)^2$. Much of the theory developed for tests based on (1.1) in normal families can be adapted to tests defined by (1.4) in multiparameter exponential famliles (Woodroofe, 1978, Lalley, 1983, Hu, 1985).

In large multi-center clinical trials it does not appear feasible to monitor the accumulating data continuously, so it is convenient to consider also "group" repeated significance

tests, in which we suppose that each "observation" z_i is actually the sum of several, say k, observations which constitute the ith group. This does not affect the theoretical developments that follow, since z_i is still normally distributed (and may be approximately normally distributed even if the individual observations are not), but it does mean that a small value of the parameter m can represent a large sample size if the group size k is large. Also, for a group sequential test with group size k the "real" expected sample size is k times the quantity in (1.3). As we shall see below, the group size k typically does not have a significant effect on the operating characteristics of a sequential test. See also Pocock (1977).

Before presenting a numerical example, it is convenient to introduce a modification of the repeated significance test defined above, which does have an important impact on its operating characteristics. As Table 1 below illustrates, a repeated significance test can have a much smaller expected sample size for large $|\mu|$ than a fixed sample test of sample size m, but the price it pays is a considerable loss of power. To recapture most of this lost power at a relatively small increase in the expected sample size, consider the following family of tests which interpolate between fixed sample tests and repeated significance tests. Given $0 < c \leq b$, let T be defined by (1.1). Stop sampling at $\min(T, m)$ and reject H_0 if either $T \leq m$ or $T > m$ and $|S_m| \geq cm^{1/2}$. The power of this test is

$$
\begin{aligned}
(1.5) \quad & P_\mu\{T \leq m\} + P_\mu\{T > m, \ |S_m| \geq cm^{1/2}\} \\
& = P_\mu\{|S_m| \geq cm\} + P_\mu\{T < m, \ |S_m| < cm^{1/2}\}.
\end{aligned}
$$

Of course the value of b must now be somewhat larger than previously if the overall significance level is to be unchanged, but by taking c only slightly larger than the rejection level of a fixed sample test, one makes the power essentially equal to the first term on the right hand side of (1.5), which in this case is about the same as for the fixed sample test. See Figure 1 and Table 2 below. This modification of the repeated significance test was suggested independently, with varying motivation, by Haybittle (1971), Peto, et al. (1976), and Siegmund (1978).

Tables 1–3 contain numerical examples. The power function has been approximated according to the suggestions in Part 2 of this paper. For comparison, results obtained by

4

Figure 1
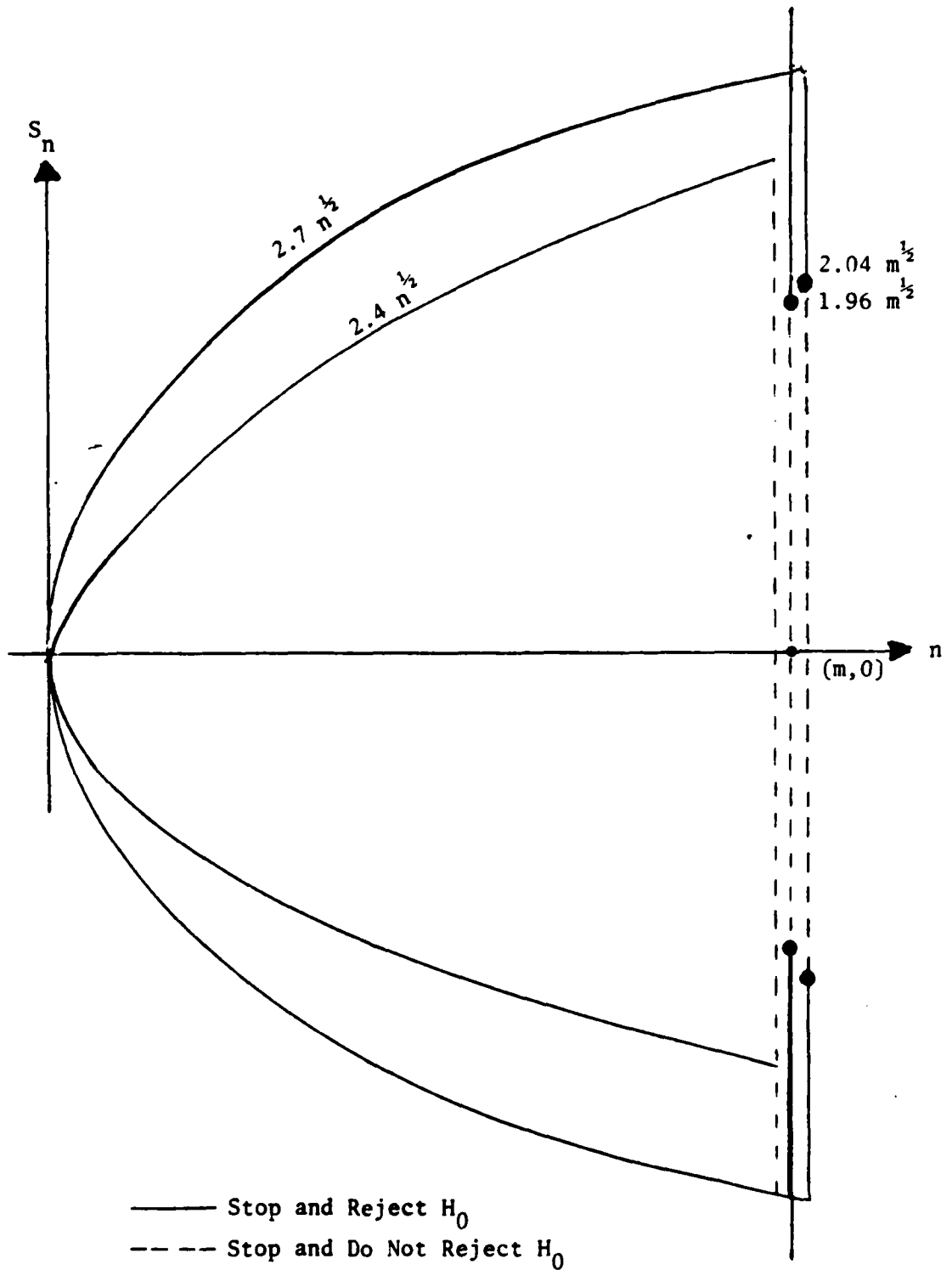
Pocock (1977) by iterative numerical integration are included in parentheses, when available. Approximations to the expected sample size are computed according to the suggestion of Siegmund (1985, equation (4.42)), which is not discussed here. Those cells which contain an asterisk are combinations of b, m, and μ for which the approximation to the expected sample size is poor. Table 1 is concerned with a repeated significance test having power function given by (1.2). It is easy to see that there is a substantial savings in the expected sample size when $|\mu| \gg 0$ compared to a fixed sample test taking m observations. To document the loss of power of the repeated significance test, the power of a fixed sample test taking m observations is also included in the table. Table 2 is concerned with a modified test having power function given by (1.5). Now there is essentially no loss of power, but still a quite considerable savings in the expected sample size. In order to compare Table 3 with Table 2, one should think of Table 2 as defining a group sequential test with $k = 10$ observations per group. Then the values given for μ in the two tables are comparable (i.e. a value in Table 3 equals the corresponding value in Table 2 divided by $k^{1/2} = 3.16$); and the expected sample sizes are comparable if one multiplies the entries in Table 2 by $k = 10$. To the accuracy of the approximations used, the group test has the same power function and just a slightly larger expected sample size than the test which inspects the data continuously.

Table 1

Repeated Significance Test

$b = 2.413$, $\quad m_0 = 1$, $\quad m = 5$, $\quad \alpha \cong .049$ (.05)

μ	Power (1.2)	$E_\mu(T \wedge m)$	Power of Fixed Sample Test
2.071	.99 (.99)	1.93 (2.05)	1.00
1.759	.95 (.95)	2.43 (2.53)	.98
1.592	.91 (.90)	2.76 (2.84)	.95
1.311	.76 (.75)	3.35 (3.41)	.83
.994	.52 (.50)	4.02 (4.04)	.60

Parenthetical entries from Pocock (1977)

Table 2

Modified Repeated Significance Test

$b = 2.7$, $c = 2.04$, $m_0 = 1$, $m = 5$, $\alpha \cong .0504$

μ	Power (1.5)	$E_\mu(T \wedge m)$	$P_\mu\{T \leq m\}$
2.071	1.00	2.26	.98
1.759	.97	2.83	.91
1.592	.94	3.18	.84
1.311	.82	3.76	.66
.994	.59	*	.40

Table 3

Modified Repeated Significance Test

$b = 2.91$, $c = 2.05$, $m_0 = 10$, $m = 50$, $\alpha \cong .0503$

μ	Power	$E_\mu(T \wedge m)$	$P_\mu\{T \leq m\}$
.655	1.00	19	.97
.556	.97	25	.89
.503	.94	29	.81
.415	.82	35	.62
.314	.59	*	.36

Remark 1.6. It is easy to devise other tests which behave about the same as the modified repeated significance test discussed here. One possibility, suggested independently by Miller (1970), Samuel-Cahn (1974), and O'Brien and Fleming (1979), is to stop at $\min(N, m)$, where $N = \min\{n : |S_n| \geq B\}$, and to reject $H_0 : \mu = 0$ if $N \leq m$. While the properties of these tests are similar to those of the modified repeated significance tests defined above, they appear to have some disadvantages. For example, their overall significance level is more sensitive to the choice of m, which makes them less flexible in adjusting to an unanticipated change in the maximum sample size once an experiment has begun. See Siegmund (1985).

A modified repeated significance test is designed to produce a fixed sample size m unless there is a substantial treatment effect as measured by the parameter of primary

interest, μ. We assume that if $|\mu|$ is large, the preference for one treatment is so strong that other considerations are essentially irrelevant. However, there typically are other measures of treatment effect which one wants to explore, especially if $\mu \cong 0$; but because of their secondary importance they do not enter into the definition of the stopping rule. There are undoubtedly also cases where if μ is close to 0, one would like to terminate the experiment as soon as possible because of economic considerations.

One can easily obtain reasonable tests which provide for early termination when H_0 appears to be true by splicing together "one-sided" tests. For example, we consider initially a modified repeated significance test of $H_0 : \mu = 0$ against $H_1 : \mu > 0$ defined by the stopping time

$$T_1 = \inf\{n : n \geq m_{01}, \ S_n \geq b_1 n^{1/2}\}$$

with rejection of H_0 if $T_1 \leq m$ or $T_1 \geq m$ and $S_m \geq cm^{1/2}$. Now consider adding a lower stopping boundary

$$(1.7) \qquad T_2 = \inf\{n : n \geq m_{02}, \ S_n \leq -b_2 n^{1/2} + \delta n\} \qquad (\delta > 0),$$

and define a new test which stops sampling at $T_1 \wedge T_2 \wedge m$ with rejection of H_0 if $T_1 \leq T_2 \wedge m$ or $T_1 \wedge T_2 > m$ and $S_m \geq cm^{1/2}$. (Here we are assuming that $-b_2 m^{1/2} + \delta m \leq cm^{1/2}$.) Presumably δ is chosen to be a positive treatment effect which it is important to detect. Since one hopes to accomplish something different with T_2 than with T_1 there is no obvious reason that the lower boundary should have the same shape as the upper boundary, or if it has, that b_1 and b_2 should have any particular relation. Nevertheless, for the convenience of this theoretical discussion, we assume that $m_{01} = m_{02} = m_0$ and $b_1 = b_2 = b$, say.

The power of this test is

$$(1.8) \qquad P_\mu\{T_1 \leq T_2 \wedge m\} + P_\mu\{T_1 \wedge T_2 > m, \ S_m \geq cm^{1/2}\},$$

which is difficult to compute exactly, but which usually is easily approximated by results developed to deal with (1.5). One approximation to (1.8) is

$$(1.9) \qquad P_\mu\{S_m \geq cm^{1/2}\} + P_\mu\{T_1 < m, \ S_m < cm^{1/2}\} - P_\mu\{T_2 < m, \ S_m \geq cm^{1/2}\}.$$

It may be shown that the difference between (1.8) and (1.9) is

$$P_\mu\{T_1 < T_2 < m, \; S_m \geq cm^{1/2}\} - P_\mu\{T_2 < T_1 < m, \; S_m < cm^{1/2}\},$$

which involves sample paths which first cross one stopping boundary, then the other, and have partially crossed the continuation region again by time m. These probabilities are usually insignificantly small unless m is close to the point where the upper and lower boundaries meet, in which case m can probably be reduced without adversely affecting the overall properties of the test (cf. Anderson, 1960). For the somewhat simpler case of a truncated sequential probability ratio test, Siegmund (1985, III.6) shows that the corresponding approximation is a good one.

For a numerical example consider the test of Table 2, which has a significance level of about .025 as a one-sided test against $H_1 : \mu > 0$. If we now introduce a lower stopping boundary (1.7) with $\delta = 1.759$ and decrease c slightly to 2.02, the approximation (1.9) indicates that the significance level of the new test is again about .025 and the power at $\mu = 1.759$ is still .97. At $\mu = .994$ the power is about .58, so introduction of the lower stopping boundary appears to lead to a negligible loss of power. On the other hand, the expected sample size when $\mu = 0$ is roughly the same as the expected sample size in Table 2 for $\mu = 1.759$, or about 2.8. This is a considerable reduction from the expected sample size of a repeated significance test, which is just slightly less than the maximum sample size, $m = 5$.

In recent years various authors have attempted to define attained significance levels, or p-values, and confidence sets relative to sequential tests. Both of these concepts require that the possible outcomes be ordered so that one knows what it means to say that one outcome is more extreme than another. For example, suppose that we use the stopping rule (1.1) and the test terminates at $T = n \in (m_0, m]$. It seems reasonable to say that a more extreme result would be a sample outcome which terminates at this or a smaller value of T and hence to define the (two-sided) attained significance level of the observed result to be $P_0\{T \leq n\}$. By similar reasoning, one can define a confidence interval for μ. For a lower $(1 - \alpha)\,100\%$ confidence bound, if $T = n \in (m_0, m]$ and $S_T > 0$, we can take for a bound

that value μ_* which satisfies

$$P_{\mu_*}\{T \le n, \, S_T > 0\} = \alpha.$$

The bound is defined similarly if $T = m_0$, $T > m$, etc.

For $b = 2.413$ as in Table 1, the attained significance of $T = 2$ according to the preceding definition is about .027. Thus, in spite of the dramatic action of stopping the test after 40% of its projected duration the evidence against H_0 as measured by the p-value is by no means dramatic. The situation is somewhat different for a modified repeated significance test if b is taken sufficiently large. For $b = 2.71$ as in Table 2, $P_0\{T \le 2\} \cong .012$; and the attained significance of any result which terminates the test before time $m = 5$ is bounded by $P_0\{T \le 5\} \cong .023$. Of course, it would defeat the purpose of using a sequential test if one insisted that the p-value be extremely small before stopping the experiment.

It seems difficult to give a persuasive theoretical justification for the definitions suggested here, and hence the principal argument in support of them is that several authors have independently arrived at essentially the same definitions. Berk and Brown (1978) discuss different alternatives. One is to order sample outcomes according to the value of $S_{T\wedge m}/T \wedge m$. If one neglects excess over the stopping boundary, this definition is equivalent to the one suggested above. However, it has the advantage that it generalizes directly to the case in which additional data become available after the experiment has terminated. Usually these data are a small part of the total sample and have an insignificant effect on the analysis. See Samuel-Cahn and Wax (1985) for an interesting example to the contrary. Siegmund (1985) contains additional references and a more complete discussion.

1.2 Sequential Survival Analysis

The discussion of the preceding section is extremely simplified, and to see how it provides considerable insight for more realistic models, we consider next the possibility of using sequential methods in clinical trials involving survival data, analyzed by the proportional hazards model (Cox, 1972). The notation is unavoidably complicated.

Suppose that patients arrive (are born) at times $y_1, y_2, \cdots$. Associated with the ith patient is a triple (z_i, z_i, c_i), where z_i is a covariate, z_i is the length of survival (age at death),

and c_i denotes the time of censoring after arrival. The assumption of the proportional hazards model is that

$$P\{z_i \in [s, s + ds) \mid z_i, z_i \geq s\} = d\Lambda_i(s) = \exp(\beta z_i)\lambda(s)ds,$$

for some unknown parameter β and base line hazard function λ. Also let $R(t, s) = \{i : y_i \leq t - s, \ z_i \wedge c_i \geq s\}$ denote those patients who are at risk at time t and whose age (measured from arrival) is at least s. Let

$$N_i(t, s) = I\{y_i + z_i \leq t, \ z_i \leq c_i, \ z_i \leq s\}$$

be the indicator that the ith patient arrived and died before time t, died at an age $\leq s$, and was not censored at the epoch of death. Cox (1972, 1975) suggested that this model be analyzed by applying likelihood methods to the log "partial" likelihood function

$$\ell(t, \beta) = \sum_i \int_{[0,t]} \left\{ \beta z_i - \log\left[\sum_{j \in R(t,s)} \exp(\beta z_j) \right] \right\} N_i(t, ds).$$

In particular, consider the score process $\dot{\ell}(t, \beta), = \partial \ell(t, \beta)/\partial \beta$, or more generally, the two parameter process

(1.10)
$$\dot{\ell}(t, s, \beta) = \sum_i \int_{[0,s]} \{z_i - \mu_\beta(t, u)\} N_i(t, du)$$

with

$$\mu_\beta(t, u) = \frac{\sum_{j \in R(t,u)} z_j \exp(\beta z_j)}{\sum_{j \in R(t,u)} \exp(\beta z_j)}.$$

It is easy to see that $\dot{\ell}(t, \beta) = \dot{\ell}(t, t, \beta)$. The score process can be used directly to test the hypothesis $H_0 : \beta = \beta_0$, and its zeroes yield partial maximum likelihood estimators of β. The asymptotic distribution theory used for probability calculations is based on the fact that under mild regularity conditions

$$\dot{\ell}(t, \beta)/\{-\ddot{\ell}(t, \beta)\}^{1/2}$$

has asymptotically a standard normal distribution. (Here $\ddot{\ell}(t, \beta)$ is the second derivative of the log partial likelihood with respect to β.) See Cox (1975) for an informal treatment and Gill (1980) for a sophisticated discussion based on martingale theory.

11

The appropriate generalization for purposes of sequential analysis is that $\dot{\ell}(t,\beta)$, when plotted against $-\ddot{\ell}(t,\beta)$ as the "time" parameter, behaves like standard Brownian motion. By virtue of the Taylor series expansion

$$\dot{\ell}(t,\beta_0) = \dot{\ell}(t,\beta) + (\beta - \beta_0)\{-\ddot{\ell}(t,\beta_0)\} + o(\beta - \beta_0),$$

we see that for β close to β_0, the test statistic $\dot{\ell}(t,\beta_0)$ plotted against $\{-\ddot{\ell}(t,\beta_0)\}$ as time behaves like Brownian motion with drift $\beta - \beta_0$ when β is the true value of the parameter. See Figure 2.
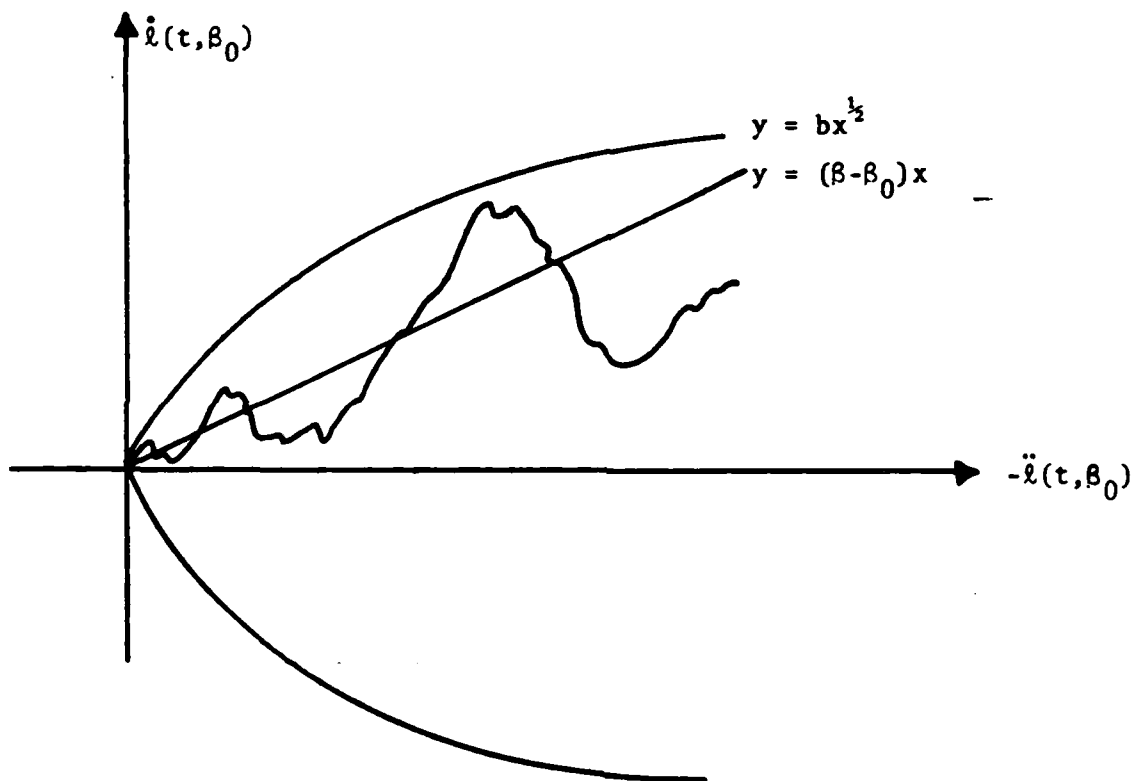

Figure 2

Formulating precisely and proving the claims of the preceding paragraph are a substantial undertaking, which is not attempted here. Sellke and Siegmund (1983) give a fairly complete discussion under the additional assumption that the triples (z_i, z_i, c_i) are indepen-

dently and identically distributed. A still more difficult argument is required if, as seems desirable, one treats the z's and c's as ancillary and conditions on their values (Sellke, 1985).

The reason that it is much more difficult to study the score function as a process in t than marginally for fixed t is briefly the following. By rewriting (1.10) as

$$\dot{\ell}(t, s, \beta) = \sum_i \int_{[0,s]} \{z_i - \mu_\beta(t, u)\}[N_i(t, du) - I\{i \in R(t, u)\}dA_i(u)],$$

one can easily show that (1.10) is a martingale in s for each fixed t. Hence martingale central limit theory is tailor-made to study the behavior of (1.10) as a process in s and in particular its marginal distribution for $s = t$. However, (1.10) with $s = t$ is not in general a martingale in t (although it is in the degenerate case that all arrival times y_i are the same). Sellke and Siegmund (1983) show that $\dot{\ell}(t, t, \beta)$ can, however, be approximated by a martingale uniformly in t; and they then apply martingale central limit theory to this approximating martingale. Sellke (1985) observes and exploits the fact that for $t_1 < t_2 < t_3 < t_4$

$$\dot{\ell}(t_4, s, \beta) - \dot{\ell}(t_3, s, \beta) \quad \text{and} \quad \dot{\ell}(t_2, s, \beta) - \dot{\ell}(t_1, s, \beta)$$

are orthogonal martingales in s.

It is customary for data monitoring committees to meet at roughly equal intervals of real time (e.g. every six months). According to the central limit theory discussed above, if one proposed to apply the methods developed in the preceding section, intervals of time should not be measured by equal calendar intervals, but by equal intervals of increase in the observed Fisher information, $\{-\ddot{\ell}(t, \beta)\}$. Siegmund (1985) describes a Monte Carlo experiment, which among other things indicates that this discrepancy may not be important – at least if the arrival and censoring mechanisms are not too erratic.

1.3 Example.

A sequential clinical trial which has recently been described in considerable detail in the medical-statistical literature is the randomized trial of propranolol conducted by the β-Blocker Heart Attack Trial Research Group (cf. BHAT, 1982, and DeMets, Hardy, Friedman, and Lan, 1984). Over a period of about twenty-seven months 3837 victims of acute myocardial infarction were randomized to a placebo group (1921) or a treatment

13

group (1916). The principal endpoint was a survival time, which was assumed to follow a proportional hazards model.

The data monitoring committee planned reviews of the results to date at $t = 1, 1.5, 2, 2.5, 3, 3.5,$ and 4 years. It was assumed that these would correspond to seven reviews at approximately equally spaced increments of increase in the observed Fisher information. The stopping rule used in this experiment was defined by parallel straight lines as in Remark 1.6, but for illustrating the theory developed here, we shall consider a repeated significance test. (For a discussion of the various factors in addition to the "stopping rule" which went into the actual decision to terminate the experiment, see DeMets et al., 1984.)

Let t_n denote the time of the nth planned inspection, $n = 1, 2, \cdots, 7$, and consider the stopping rule

$$(1.11) \qquad T = \inf\{t_n : n \geq m_0, \; |\dot{\ell}(t_n, 0)| \geq b[-\ddot{\ell}(t_n, 0)]^{1/2}\}.$$

To test the hypothesis $H_0 : \beta = 0$ of no treatment effect, stop sampling at $\min(T, t_7)$ and reject H_0 if either $T \leq t_7$ or $T > t_7$ and $|\dot{\ell}(t_7, 0)| \geq c[-\ddot{\ell}(t_7, 0]^{1/2}$. The normal approximation described above indicates that $m_0 = 2$, $m = 7$, $b = 2.65$, and $c = 2.05$ yield a .05 level test having a power function very close to that of the sequential design used in BHAT (1982). The power function and approximate expected sample size for this test in the approximating normal model are given in Table 4.

Table 4

Repeated Significance Test for Normal Data

$b = 2.65, \; c = 2.05, \; m_0 = 2, \; m = 7$		
μ	Power	Expected Sample Size
1.50	.97	3.59
1.25	.90	4.47
.75	.50	*
.00	.05	*

To relate the power as a function of the normal mean μ to the parameter β, it is

necessary to make some assumptions about the rate of increase of $-\ddot{\ell}(t,0)$. For the simple model we have discussed, if the censoring mechanism does not depend on the covariate, it is easy to see that for β close to 0 each death yields on the average about $\frac{1}{4}$ unit of information. In the 3.5 years before this experiment was terminated there were 318 deaths for an accumulation of approximately 79.5 units of information, or an average of 13.3 units per inspection period. This means that a value of μ in Table 4 corresponds roughly to a value of $\beta = \mu/(13.3)^{1/2}$. In particular the row for $\mu = 1.25$ in Table 4 corresponds to β about equal to .34. (The discussion of sample size selection in BHAT, 1981, shows the expectation before the experiment began of a somewhat more rapid rate of accumulation of information, hence greater power, than actually occurred.)

Similarly, the expected sample size in Table 4 multiplied by an average rate of accumulation of information gives the expected information until termination of the experiment. This in itself may not be as meaningful as, for example, the expected number of deaths or the expected real time of the experiment. If we use the approximation that information equals one fourth the number of deaths, then expected information is directly proportional to a more meaningful quantity. Without much stronger modeling assumptions, involving the arrival rate and the baseline hazard function, there is no relation between expected information and expected real time for the experiment. Qualitatively, information accumulates more slowly early in the experiment, so a reduction in expected information of 50% compared to, say, a fixed sample test, invariably means a smaller reduction in the expected time of the experiment.

The observed values of $\dot{\ell}/(-\ddot{\ell})^{1/2}$ at $1, 1.5, \cdots, 3.5$ years were respectively 1.68, 2.24, 2.37, 2.30, 2.34, 2.82 (DeMets, et al., 1984). For the test actually used and also for the repeated significance test suggested above, these data lead to termination of the experiment at $t = 3.5$ years, or six months before the final planned inspection. (A more detailed analysis using a number of covariates gave the value 3.05 for the corresponding normalized statistic, which is reasonably consistent with the 2.82 of the simplest possible model. See BHAT, 1982.) The values of $\dot{\ell}(t,0)$ and $-\ddot{\ell}(t,0)$ are not given separately, so it is not possible to plot $\dot{\ell}$ against $-\ddot{\ell}$ as in Figure 2. This is unfortunate because such a plot would allow one to check whether inspections indeed occur at approximately equal increments of increase of

information; and much more importantly, since the plot should be approximately a straight line with slope β, it would give a visual estimate of β and a visual goodness of fit test of the proportional hazards model. See Siegmund (1985) for a hypothetical reconstruction of $-\tilde{\ell}$ based on the assumption that each death contributes one fourth unit of information and calculation of a confidence interval for β.

According to the definition of attained significance given in Section 1.1, the two-sided p-value of $T = 6$ in the approximating normal model is $P_0\{T \leq 6\} \cong .023$. The attained level announced in BHAT (1982) is .01, which seems to be too small – even if one takes into account the different stopping rule and the possibility of slight variations in the definition of the p-value.

2. Boundary Crossing Probabilities

2.1 Introduction and Asymptotic Normalisation

In this section we consider the mathematical problem of calculating approximately probabilities like (1.5).

Let $z_1, z_2, \cdots$ be independent and identically distributed, and set $S_n = z_1 + \cdots + z_n$. Let $c(n)$, $n = 1, 2, \cdots$ be constants and $m_0 < m$ positive integers. Define the stopping time

$$T = \inf\{n : n \geq m_0, \ S_n \geq c(n)\},$$

and consider the problem of evaluating

(2.1) $$P\{T \leq m\}$$

or

(2.2) $$P\{T < m \mid S_m = \xi\}.$$

Since $P\{T \leq m\} = P\{S_m \geq c(m)\} + \int_{-\infty}^{c(m)} P\{T < m \mid S_m = \xi\}P\{S_m \in d\xi\}$, and since the distribution of S_m is comparatively easy to evaluate, at least approximately, a good approximation to (2.2) usually yields a good approximation to (2.1). Similarly, evaluation

16

of (2.2) is frequently the principle ingredient in calculating (1.5). Hence our focus in what follows is on developing approximations to (2.2), which occasionally is of interest in itself.

Since (2.2) can only rarely be evaluated exactly, it is convenient to imbed our problem in a sequence of problems and seek an asymptotic approximation. The actual calculations are preceded by some remarks about the two most obvious asymptotic formulations.

In problems scaled for large deviations, we consider the asymptotic evaluation as $m \to \infty$ of probabilities of the form

$$p(m) = P\{S_n \geq m\, c(n/m) \quad \text{for some} \quad m_0 \leq n \leq m \mid S_m = \xi\}, \quad (\xi = m\xi_0).$$

Since the boundary $m\, c(n/m)$ is $O(m^{1/2})$ standard deviations away from the (conditional) mean path of S_n, these probabilities typically converge to zero, and a reasonable approximation would be of the form $p(m) \sim q(m)$ for some easily evaluated analytic expression $q(m)$.

An alternative, the ordinary deviation or diffusion scaling, suggests consideration of

$$p'(m) = P\{S_n \geq m^{1/2}c(n/m) \quad \text{for some} \quad m_0 \leq n \leq m \mid S_m = \xi\} \quad (\xi = m^{1/2}\xi_0).$$

Now the mean path of S_n is $O(1)$ standard deviations from the boundary $m^{1/2}c(n/m)$, so typically if $m_0/m \to t_0$

$$p'(m) \to p = P\{W(t) \geq c(t) \quad \text{for some} \quad t_0 \leq t \leq 1 \mid W(1) = \xi_0\},$$

where $W(t)$, $0 \leq t < \infty$, is a Brownian motion process. The approximation of $p'(m)$ by p is often not particularly good, but it can be improved by finding an expansion of the form

$$p'(m) = p + p_1\, m^{-1/2} + o(m^{-1/2}),$$

which has been called a corrected diffusion approximation (cf. Siegmund, 1984, and references cited there).

Typically, large deviation approximations are more easily obtained than corrected diffusion approximations. This is especially true for nonlinear boundaries, $c(n)$. See Hogan (1984) for the first corrected diffusion approximations in a nonlinear case. Occasionally it is possible to write a single approximation which is applicable to both cases. When this is

so, that approximation is usually a very good one. Except for a few remarks, only large deviation scaling is considered in what follows.

Numerous methods have been invented for approximating boundary crossing probabilities (e.g. Borovkov, 1962, Woodroofe, 1976b, Lai and Siegmund, 1977, Daniels, 1974, Jennen and Lerche, 1981, Durbin, 1981). The method described below has the virtues that it is essentially the same in both discrete and continuous time, it is fairly general, and it yields exact results in most of the simple situations where exact results can be obtained. Our starting point is a derivation of the standard reflection principle for Brownian motion. The argument is then incrementally modified to deal with problems in discrete time and problems involving nonlinear boundaries. Woodroofe (1982) contains an exposition of alternative methods supported by complete proofs.

2.2 Reflection Principle for Brownian Motion

Let $W(t)$, $0 \leq t < \infty$, be Brownian motion with drift μ and unit scale parameter, and let $\mathcal{F}(t)$ denote the σ-field of events defined by $W(s)$, $0 \leq s \leq t$. It will be convenient to use the notation

$$P_\xi^{(m)}(A) = P_\mu\{A \mid W(m) = \xi\} \qquad (A \in \mathcal{F}(m)).$$

By the sufficiency of $W(m)$, this conditional probability does not depend on μ. For all $\xi_1 \neq \xi_2$ and $t < m$, the probabilities $P_{\xi_1}^{(m)}$ and $P_{\xi_2}^{(m)}$ when restricted to $\mathcal{F}(t)$ are mutually absolutely continuous; and a straightforward calculation shows that the likelihood ratio of $W(s)$, $s \leq t$, under $P_{\xi_1}^{(m)}$ relative to $P_{\xi_2}^{(m)}$ is

$$(2.3) \qquad \ell^{(m)}(t, W(t); \, \xi_1, \xi_2) = \exp\left\{\left[(\xi_1 - \xi_2)W(t) - \frac{t}{2m}(\xi_0^2 - \xi_1^2)\right] / (m - t)\right\}.$$

The following is a version of Wald's likelihood ratio identity, which can be proved by standard martingale arguments.

Proposition 2.4. For any $\xi_1 \neq \xi_2$, $m > 0$, stopping time T and event $A \in \mathcal{F}(T)$

$$P_{\xi_1}^{(m)}(A \cap \{T < m\}) = E_{\xi_2}^{(m)}[\ell^{(m)}(T, W(T); \, \xi_1, \xi_2); \, A \cap \{T < m\}],$$

where $\ell^{(m)}$ is given by (2.3).

Let $b > 0$, $-\infty < \eta < \infty$, and define

$$\tau = \inf\{t : W(t) \geq b + \eta t\}.$$

Let $\xi_1 = \xi < b + \eta m$, and let $\xi_2 = 2(b + \eta m) - \xi$ be ξ "reflected" about $b + \eta m$. See Figure 3. Since $W(\tau) = b + \eta\tau$ on $\{\tau < m\}$, and $P_{\xi_2}^{(m)}\{\tau < m\} = 1$, from Proposition 2.4 and simple algebra one obtains the well-known result

$$(2.5) \qquad P_\xi^{(m)}\{\tau < m\} = \exp[-2b(b + \eta m - \xi)/m].$$

Siegmund and Yuh (1982) show how a slightly more sophisticated version of this argument yields Anderson's (1960) results.

2.3 Correction for Discrete Time

Consider now the same problem in discrete time, so

$$\tau = \inf\{n : S_n \geq b + \eta n\},$$

where $S_n = z_1 + \cdots + z_n$, and under P_μ the z's are independent normally distributed random variables with mean μ and variance 1. Now the preceding argument yields

$$
\begin{aligned}
(2.6) \qquad & P_\xi^{(m)}\{\tau < m\}\exp[2b(b + \eta m - \xi)/m] \\
& = E_{\xi_2}^{(m)}[\exp\{-2(b + \eta m - \xi)[S_\tau - b - \eta\tau]/(m - \tau)\};\ \tau < m],
\end{aligned}
$$

where $\xi_2 = 2(b + \eta m) - \xi$.

To analyze the right hand side of (2.6) asymptotically, suppose that $b = \varsigma m$ and $\xi = m\xi_0$ for some fixed $\varsigma > 0$ and $\xi_0 < \varsigma + \eta$. Since the $P_{\xi_2}^{(m)}$-deviations of S_n from its expectations, $[2(\varsigma + \eta) - \xi_0]n$, are of order $n^{1/2}$, a law of large numbers argument shows that with probability approaching 1, S_n crosses the line $\varsigma m + \eta n$ near where its line of drift does, so for any $\epsilon > 0$

$$(2.7) \qquad \lim_{m \to \infty} P_{\xi_2}^{(m)}\{|m^{-1}\tau - \varsigma(2\varsigma + \eta - \xi_0)^{-1}| > \epsilon\} = 0.$$

See Figure 3. It follows that the right hand side of (2.6) has the same asymptotic behavior as

$$E_{\xi_2}^{(m)}[\exp\{-2(2\varsigma + \eta - \xi_0)(S_\tau - \varsigma m - \eta\tau);\ \tau < m].$$
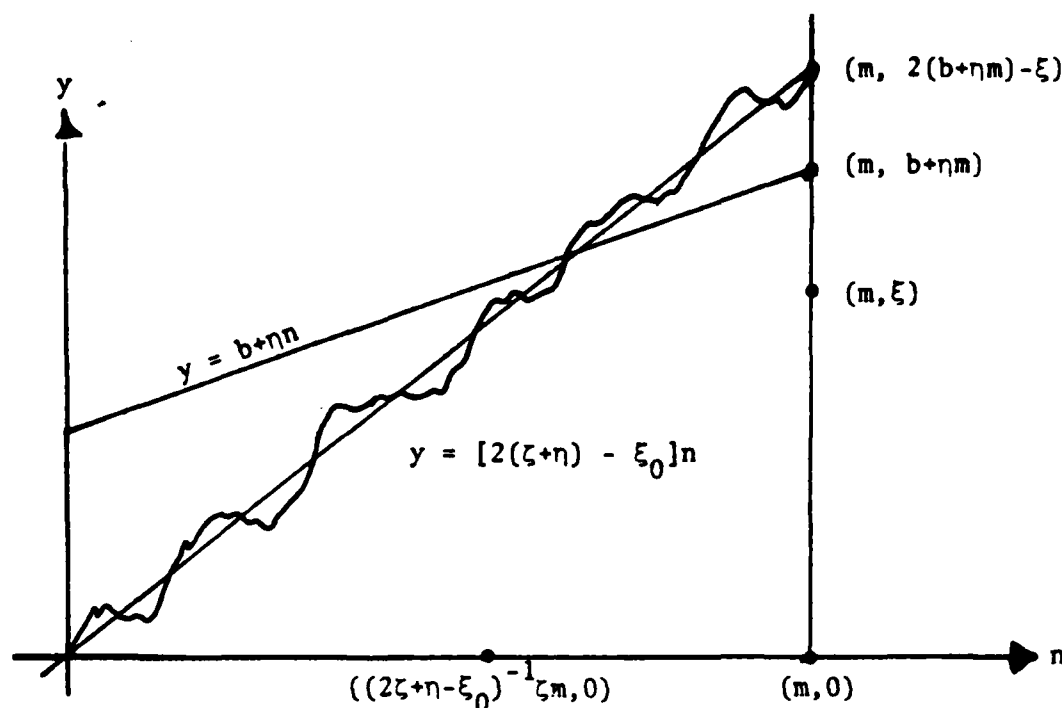
Figure 3

If this expectation were with respect to the unconditional probability with the same drift, $P_{2(\zeta+\eta)-\xi_0}$, one could apply the renewal theorem in the manner which Feller (1972, Chapter XII) uses to derive Cramér's estimate for the probability of ruin, and hence evaluate (2.8) in the limit as $m \to \infty$. Specifically, observe that for a random walk $\tilde{S}_n$, $n = 1, 2, \cdots$, with non-negative drift $\tilde{\mu} = E\tilde{S}_1$ and for

$$\tilde{\tau} = \inf\{n : \tilde{S}_n \geq a\},$$

$\tilde{S}_{\tilde{\tau}} - a$ can be regarded as the residual lifetime in a renewal process defined by $\tilde{S}_{\tilde{\tau}_+}$, where $\tilde{\tau}_+ = \inf\{n : \tilde{S}_n > 0\}$. Hence if $\tilde{S}_1$ is nonarithmetic the renewal theorem implies that as $a \to \infty$

$$(2.8) \qquad P\{\tilde{S}_{\tilde{\tau}} - a \leq x\} \to [E(\tilde{S}_{\tilde{\tau}_+})]^{-1} \int_0^x P\{\tilde{S}_{\tilde{\tau}_+} \geq y\} dy.$$

See Feller (1972, Chapters XI and XII). For a discussion which is oriented towards the present application, see Siegmund (1985, Chapter VIII).

During the relatively short time interval in which according to (2.7) τ falls with proba-

bility close to one, the increments to the conditional $P_{\xi_2}^{(m)}$ process S_n and the unconditional $P_{2\varsigma+\eta-\xi_0}$ process both behave essentially the same, so the $P_{\xi_2}^{(m)}$ and $P_{2(\varsigma+\eta)-\xi_0}$ limiting distributions of $S_r - \varsigma m - \eta r$ are the same, and are given by (2.8) with $\tilde{S}_n = S_n - \eta n$.

One simple way to make this argument precise is to obtain a slightly different version of (2.6) by using Wald's likelihood ratio identity to differentiate $P_{\xi}^{(m)}$ with respect to $P_{2(\varsigma+\eta)-\xi_0}$ instead of $P_{\xi_2}^{(m)}$. Let $\mu_2 = \xi_2/m = 2(\varsigma+\eta)-\xi_0$. An easy calculation shows that the likelihood ratio of $x_1,\cdots,x_n$ under $P_{\xi_2}^{(m)}$ relative to P_{μ_2} is

$$\exp\left[-\frac{1}{2}(S_n - \mu_2 n)^2/(m-n)\right] \bigg/ \left(1 - \frac{n}{m}\right)^{1/2},$$

so the right hand side of (2.6) equals

(2.9) $$E_{\mu_2}\left[\exp\left\{-2(\varsigma+\eta-\xi_0)(S_r - b - \eta r)/\left(1 - \frac{r}{m}\right)\right.\right.$$
$$\left.\left.-\frac{1}{2}(S_r - \mu_2 r)^2/(m-r)\right\}\bigg/\left(1 - \frac{r}{m}\right)^{1/2}; \; r < m\right].$$

The asymptotic marginal distribution of the random variables appearing in (2.9) are easily determined. Under P_{μ_2}, r/m converges in probability to the same limit as in (2.7); the renewal theorem applies as in (2.8) with $\tilde{S}_n = S_n - \eta n - b$; and by an easy application of Anscombe's theorem, $(S_r - \mu_2 r)/r^{1/2}$ is asymptotically standard normal. Also by Lemma 2.16 below, $S_r - \eta r - b$ and $(S_r - \mu_2 r)/r^{1/2}$ are asymptotically independent. Thus we have all the ingredients to evaluate (2.9). From (2.8) and some calculation one sees that the limit of (2.9) is $\nu[2(2\varsigma + \eta - \xi_0)]$, where for $\mu > 0$ and $r_+ = \inf\{n : S_n > 0\}$

(2.10) $$\nu(\mu) = [1 - E_{\mu/2}\exp(-\mu\, S_{r_+})]/\mu\, E_{\mu/2}(S_{r_+}).$$

Hence by (2.6)

(2.11) $$P_{\xi}^{(m)}\{r < m\} \sim \nu[2(2\varsigma + \eta - \xi_0)]\exp[-2m\varsigma(\varsigma + \eta - \xi_0)]$$

$(b = m\varsigma, \; \xi = m\xi_0, \; \varsigma > 0, \; \xi_0 < \varsigma+\eta)$. Random walk theory (cf. Feller, 1972, Chapter XVIII or Siegmund, 1985, Chapter VIII) permits one to obtain a numerically calculable expression for (2.10), to wit

(2.12) $$\nu(\mu) = 2\mu^{-2}\exp\left[-2\sum_1^\infty n^{-1}\Phi\left(-\frac{1}{2}\mu n^{1/2}\right)\right],$$

where Φ is the standard normal distribution function. For many purposes it suffices to use the approximation

$$(2.13) \qquad \nu(\mu) = \exp(-\rho\mu) + o(\mu^2) \quad (\mu \to 0),$$

where

$$(2.14) \qquad \rho = E_0 \, S_{r_+}^2 / 2 \, E_0(S_{r_+}) = -\pi^{-1} \int_0^\infty \lambda^{-2} \log[2\lambda^{-2}\{1 - \exp(-\lambda^2/2)\}]d\lambda$$
$$\cong .583.$$

Partial justification for (2.13) comes from a Taylor series expansion of (2.10) to obtain

$$(2.15) \qquad \nu(\mu) = 1 - \mu \, E_{\mu/2}(S_{r_+}^2)/2 \, E_{\mu/2}(S_{r_+}) + \cdots.$$

This is easily turned into a proof of (2.13) with an error $o(\mu)$. That the error is actually $o(\mu^2)$ and that ρ has the value given in (2.14) are more difficult to prove. See Siegmund (1985, Chapter X) for details.

To complete the proof of (2.11), we must justify the asymptotic independence of $(S_r - \mu_2 r)/r^{1/2}$ and $S_r - \eta r - b$ used in evaluating the limit of (2.9). The first person to have noticed this relation appears to have been Stam (1968).

Lemma 2.16. Let $\tilde{S}_n$, $n = 1, 2, \cdots$ be a non-arithmetic random walk with drift $E(\tilde{S}_1) = \tilde{\mu} > 0$ and finite variance $\tilde{\sigma}^2 = \mathrm{var}(\tilde{S}_1)$. Let $\tilde{r} = \tilde{r}(a) = \inf\{n : S_n \geq a\}$. As $a \to \infty$, for all $x \geq 0$, $-\infty < y < \infty$

$$P\{\tilde{S}_{\tilde{r}} - a \leq x, \ (\tilde{r} - a\tilde{\mu}^{-1})/(a\tilde{\sigma}^2\tilde{\mu}^{-3})^{1/2} \leq y\} \to H(x)\Phi(y)$$

and

$$P\{\tilde{S}_{\tilde{r}} - a \leq x, \ (\tilde{S}_{\tilde{r}} - \tilde{\mu}\tilde{r})/\tilde{r}^{1/2} \leq y\} \to H(x)\Phi(y),$$

where H is the distribution function given in (2.8).

Remark. A similar result holds for arithmetic random walks, but the distribution H is slightly different. A result corresponding to the first relation in Lemma 2.16 holds if $\tilde{\mu} = 0$, but in this case the appropriately normalized $\tilde{r}$ is $\tilde{r}/a^2$, and Φ must be replaced by $2\Phi(y^{-1/2}) - 1$.

Proof of Lemma 2.16. Since by (2.8)

$$(\tilde{S}_{\tilde{\tau}} - \tilde{\mu}\tilde{\tau})/\tilde{\tau}^{1/2} - (a - \tilde{\mu}\tilde{\tau})/\tilde{\tau}^{1/2} \xrightarrow{P} 0,$$

the second asymptotic relation follows from the first one. To prove the first one, let $n = n(a, y) = a\tilde{\mu}^{-1} + y(a\tilde{\sigma}^2\tilde{\mu}^{-3})^{1/2}$. Then

$$P\{\tilde{S}_{\tilde{\tau}} - a \le x, \ \tilde{\tau} > n\} = E[P\{\tilde{S}_{\tilde{\tau}} - a \le x \mid \tilde{\tau} > n, \ \tilde{S}_n\}; \tilde{\tau} > n].$$

Suppose $a_1 < a$, $a - a_1 \to \infty$, but $a - a_1 = o(a^{1/2})$. Then by the central limit theorem

$$E[P\{\tilde{S}_{\tilde{\tau}} - a \le x \mid \tilde{\tau} > n, \ \tilde{S}\}; \tilde{\tau} > n, \ a_1 \le \tilde{S}_n < a] \le P\{a_1 \le \tilde{S}_n < a\} \to 0.$$

Also, uniformly on $\{\tilde{\tau} > n, \ \tilde{S}_n < a_1\}$, by (2.8)

$$P\{\tilde{S}_{\tilde{\tau}(a)} - a \le x \mid \tilde{\tau} > n, \ \tilde{S}_n = z\} = P\{\tilde{S}_{\tilde{\tau}(a-z)} - (a - z) \le x\} \to H(x).$$

Hence

$$P\{\tilde{S}_{\tilde{\tau}} - a \le x, \ \tilde{\tau} > n\} = H(x)P\{\tilde{\tau} > n, \ \tilde{S}_n < a_1\} + o(1)$$

$$= H(x) \ P\{\tilde{\tau} > n\} + o(1).$$

The lemma follows from the well-known and easily proved asymptotic normality of $\tilde{\tau}$ with the indicated scaling.

Using (2.13), one can rewrite (2.11) in the form

$$(2.17) \qquad P_\xi^{(m)}\{\tau < m\} \cong \exp\{-2(b + \rho)(b + \rho + \eta m - \xi)/m\}.$$

This last approximation is particularly interesting because it is of the form (2.5) with b replaced by $b + \rho$. Moreover, it follows from (2.8) and (2.13)–(2.15) that ρ is approximately the expected excess of the discrete random walk over the boundary, so (2.17) has the interpretation that to correct for discrete time one can use the Brownian motion result (2.5) with boundaries displaced by the average amount the discrete time process jumps over the boundary (cf. Siegmund, 1984 for other results having a similar interpretation).

The approximation (2.17) is also valid as a corrected diffusion approximation, i.e. if $b = \varsigma m^{1/2}$, $\eta = \eta_0 m^{-1/2}$, and $\xi = \xi_0 m^{1/2}$, the difference between the two sides of (2.17) is $o(m^{-1/2})$. This result can be proved along the lines of the argument sketched above;

but the details are more difficult because the $P_{\xi_2}^{(m)}$ distribution of $m^{-1}r$ does not become degenerate as $m \to \infty$. See Siegmund (1984).

The accuracy of (2.17) is quite good. For $m = 3$, $b = 1.564$, and $\eta = 0$ Worsley (1983) has numerically calculated $P_0^{(m)}\{r < m\}$ to be .05. The approximation (2.17) yields .0463. For the corresponding comparison when $m = 5$, $b = 2.165$ ($m = 10$, $b = 3.292$), (2.17) gives .0488 (.0496). It is perhaps worth observing that an uncorrected diffusion approximation is very poor for small m - ranging from .1 to .2 for these examples.

The asymptotic relation (2.11) can be generalized to a large class of random walks whose distribution can be imbedded in an exponential family (Siegmund, 1982). The analogue of ν given in (2.10) can be computed numerically using results of Woodroofe (1979) or by an approximation along the lines of (2.13). With some technical improvements the method also works for a general class of nonlinear boundaries. The key is (2.7), which suggests that if the boundary is to be crossed at all, it will be crossed close to some distinguished point. This further suggests that one try to approximate the boundary by its tangent at this distinguished point, which can be determined as the point through which the $P_{\xi_2}^{(m)}$ line of drift passes when ξ_2 is appropriately chosen for the linear problem of the tangent line. Siegmund (1982) discusses the example of repeated significance tests in detail.

2.4 Repeated Significance Tests

For repeated significance tests in exponential families a slightly modified method requires considerably less algebraic detail. We continue to consider the case of normally distributed observations, and let T be defined by (1.1).

Theorem 2.18. Suppose $b \to \infty$, $m \to \infty$ and $m_0 \to \infty$ in such a way that for fixed $0 < \mu_1 < \mu_0 < \infty$, $bm^{-1/2} = \mu_1$, $bm_0^{-1/2} = \mu_0$. Let $0 < |\xi_0| < \mu_1$ and $\xi = m\xi_0$. Then $P_\xi^{(m)}\{T < m\} \le (\mu_0 \mu_1^{-1}) \exp[-\frac{1}{2}m(\mu_1^2 - \xi_0^2)]$; and for $\mu_1^2/\mu_0 < |\xi_0| < \mu_1$

$$P_\xi^{(m)}\{T < m\} \sim \nu(\mu_1^2/\xi_0)\mu_1\xi_0^{-1}\exp[-\frac{1}{2}m(\mu_1^2 - \xi_0^2)],$$

where ν is given by (2.12).

Corollary 2.19. Suppose that the asymptotic scaling of Theorem 2.18 holds and also

24

$cm^{-1/2} = \gamma \in (\mu_1^2/\mu_0, \mu_1)$. Then for $\mu \neq 0$

(2.20)
$$P_\mu\{T < m, |S_m| < cm^{1/2}\} \sim \frac{\varphi[m^{1/2}(\mu_1 - |\mu|)]}{|\mu| m^{1/2}} \mu_1 \gamma^{-1} \nu(\mu_1^2 \gamma^{-1}) \exp[-m|\mu|(\mu_1 - \gamma)],$$

and

(2.21)
$$P_0\{T < m, |S_m| < cm^{1/2}\} \sim 2b \, \varphi(b) \int_{\mu_1^2 \gamma^{-1}}^{\mu_0} x^{-1} \nu(x) dx,$$

where φ denotes the standard normal density function and ν is given by (2.12). The relation (2.21) also holds when $c = b$ ($\gamma = \mu_1$); (2.20) and (2.21) hold if $m_0 = o(m)$ as $b \to \infty$.

Remark. Corollary 2.19 suggests that one approximate (1.5) by using (2.20) or (2.21) as an approximation for the second term on the right hand side of (1.5). This is in fact the approximation used to compute the entries in Tables 1–3. Strictly speaking (2.20) is not a true asymptotic relation when $c = b$, but it usually gives a good approximation and is much easier to evaluate than the asymptotically "correct" result. See Siegmund (1985, IX.3) for a more complete discussion of this point.

Corollary 2.19 follows easily from the theorem, and some simple estimates which are omitted here (cf. Siegmund, 1985, IX.3). A proof of Theorem 2.18 follows.

Proof of Theorem 2.18. First observe that in the derivation of (2.5) one could pretend that time flows from m to 0 instead of from 0 to m and "reflect" the value $W(0) = 0$ to $W(0) = 2b$ instead of reflecting $W(m) = \xi$ to $W(m) = 2(b + \eta m) - \xi$. Also recall that in the derivation of (2.11) it was convenient to work with the unconditional probability with the same drift as $P_{\xi_2}^{(m)}$.

Let

$$P_{\lambda,\xi}^{(m)}(A) = P_\mu(A \mid S_0 = \lambda, \ S_m = \xi) \qquad (A \in \mathcal{F}_m),$$

and put

$$\tilde{P}_\xi^{(m)}(A) = \int_{-\infty}^{\infty} P_{\lambda,\xi}^{(m)}(A) \, \varphi[(\lambda - \xi)/m^{1/2}] d\lambda/m^{1/2}.$$

Note that if we regard the process as running backwards from an "initial" value $S_m = \xi$, then under $\tilde{P}_\xi^{(m)}$ it is normal random walk with zero drift.

Let $T^* = \sup\{n : n \le m, |S_n| \ge bn^{1/2}\}$, so

$$P_{0,\xi}^{(m)}\{T < m\} = P_{0,\xi}^{(m)}\{T^* \ge m_0\}.$$

It is easy to see that the likelihood ratio of $x_{n+1}, \cdots, x_m$ under $P_{\lambda,\xi}^{(m)}$ relative to $P_{0,\xi}^{(m)}$ is $\ell^{(m)}(m-n, S_n - \xi; \lambda - \xi, -\xi)$, where $\ell^{(m)}$ is given by (2.3). This simplifies to

$$\exp[\lambda S_n/n - \lambda^2/2n - \lambda\xi/m + \lambda^2/2m].$$

Hence by a straightforward integration one sees that the likelihood ratio of $x_{n+1}, \cdots, x_m$ under $\tilde{P}_\xi^{(m)}$ relative to $P_{0,\xi}^{(m)}$ is

(2.22)
$$\int_{-\infty}^{\infty} \exp[\lambda S_n/n - \lambda^2/2n - \lambda\xi/m + \lambda^2/2m]\varphi[(\lambda - \xi)/m^{1/2}]d\lambda/m^{1/2}$$
$$= (n/m)^{1/2}\exp[S_n^2/2n - \xi^2/2m].$$

Since T^* is a stopping time for the process running backwards from time m to time 0, Wald's likelihood ratio identity yields the representation

(2.23)
$$P_{0,\xi}^{(m)}\{T < m\} = P_{0,\xi}^{(m)}\{T^* \ge m_0\}$$
$$= \tilde{E}_\xi^{(m)}\{(m/T^*)^{1/2}\exp[-S_{T^*}^2/2T^* + \xi^2/2m]; \; T^* \ge m_0\}$$
$$= \exp[-\frac{1}{2}(b^2 - \xi^2/m)]\tilde{E}_\xi^{(m)}\{(m/T^*)^{1/2}\exp[-\frac{1}{2}(S_{T^*}^2/T^* - b^2)]; \; T^* \ge m_0\},$$

where $\tilde{E}_\xi^{(m)}$ denotes expectation with respect to $\tilde{P}_\xi^{(m)}$.

The inequality in Theorem 2.18 is an immediate consequence of (2.23). To prove the asymptotic relation, it remains to evaluate asymptotically the expectation in (2.23).

Observe that the $\tilde{P}_\xi^{(m)}$ joint distribution of (T^*, S_{T^*}) is the same as the P_0 joint distribution of $(m - r^*, \xi + S_{r^*})$, where

$$r^* = \inf\{n : |\xi + S_n| \ge b(m - n)^{1/2}\}.$$

Hence the expectation on the right hand side of (2.23) equals

(2.24)
$$E_0\{(1 - r^*/m)^{-1/2}\exp[-\frac{1}{2}(\xi + S_{r^*})^2/(m - r^*) - b^2)]\}; \; r^* \le m - m_0\}.$$

An easy law of large numbers argument shows that as $m \to \infty$

(2.25)
$$m^{-1}r^* \to 1 - (\xi_0/\mu_1)^2$$

in probability, and in particular

$$(2.26) \qquad P_0\{r^* \le m - m_0\} \to \begin{cases} 1 & \text{for } |\xi_0| > \mu_1^2/\mu_0 \\ 0 & \text{for } |\xi_0| < \mu_1^2/\mu_0. \end{cases}$$

If we were dealing with Brownian motion, for which there would be no excess over the boundary, this would complete the argument. For the discrete time process, after using (2.25) and (2.26) in (2.24), it suffices to show that

$$(2.27) \qquad \lim_{m \to \infty} E_0 \left\{ \exp\left[-\frac{1}{2}(\xi + S_{r^*})^2/(m - r^*) - \mu_1^2 m \right] \right\} = \nu(\mu_1^2/\xi_0),$$

which requires a renewal theorem (cf. (2.8)) for nonlinear functions of a random walk.

To verify (2.27) observe that r^* can be expressed

$$(2.28) \qquad r^* = \inf\left\{ n : n \ge 1, \ \frac{1}{2}\mu_1^2 \xi_0^{-1} n + S_n + \frac{1}{2}m^{-1}\xi_0^{-1}S_n^2 \ge \frac{1}{2}\xi_0^{-1}m(\mu_1^2 - \xi_0^2) \right\}.$$

If the term involving S_n^2 did not appear in this expression, the renewal theorem would give us the limiting distribution of the excess over the boundary. Because of (2.25) It seems plausible that in the relatively small interval of time into which r^* falls with probability close to one, the quadratic term $m^{-1}S_n^2$ is effectively constant and hence has no effect on this limiting distribution. Lai and Siegmund (1977) describe a general class of processes which can be decomposed into the sum of a random walk and a term which varies sufficiently slowly that the limiting distribution of excess over the boundary is determined by the random walk alone via (2.8). Lai and Siegmund's result is not directly applicable here, but their method is. See Appendix 2 for an informal discussion of nonlinear renewal theory. The consequence is that

$$\lim_{m \to \infty} P_0 \left\{ \frac{1}{2}\mu_1^2 \xi_0^{-1} r^* + S_{r^*} + \frac{1}{2}m^{-1}\xi_0^{-1}S_{r^*}^2 - \frac{1}{2}\xi_0^{-1}m(\mu_1^2 - \xi_0^2) \le z \right\} = H(z),$$

where $H(z)$ is the limiting distribution as given in (2.8) for a random walk $\tilde{S}_n$ having normally distributed increments with mean $\frac{1}{2}\mu_1^2\xi_0^{-1}$ and variance 1. With the aid of (2.25) it is easy to convert this limiting result to

$$\lim_{m} P_0 \left\{ \frac{1}{2}(m - r^*)^{-1}(m\xi_0 + S_{r^*})^2 - \frac{1}{2}\mu_1^2 m \le z \right\} = H(z\xi_0\mu_1^{-2}).$$

27

A trivial change of variable yields (2.27) with the same function ν that appears in (2.11) as the limit of (2.9).

The method described above generalizes in a straightforward fashion to repeated significance tests in one parameter exponential families. For the much more difficult multiparameter case, see Woodroofe (1978) and Lalley (1983). Hu (1985) shows that the present method leads to simplifications and new results in the multiparameter case, especially when there is some invariance present.

In the case of Brownian motion the preceding argument can easily be sharpened to yield a second order term in an asymptotic expansion of $P_{\xi}^{(m)}\{T < m\}$ or $P_{\mu}\{T < m,\ |W(m)| < cm^{1/2}\}$. When there is no excess over the boundary the only approximation involved in the preceding argument is that of replacing $(1 - \tau^{*}/m)^{-1/2}$ by its limit as given by (2.25). To obtain the next order of approximation, it is only necessary to expand $(1 - \tau^{*}/m)^{-1/2}$ in a Taylor series and analyze its central limit behavior. Although simple in principle, the calculation is quite complicated in detail because one must consider three cases: ξ_0 close to the endpoints of the interval $(\mu_1^2/\mu_0,\ \mu_1)$ and ξ_0 in the interior of this interval. Siegmund (1985) shows under the conditions of Corollary 2.19, for T defined by (1.1) with Brownian motion $W(t)$ instead of S_n, for $\mu_1^2/\mu_0 < \gamma \leq \mu_1$

(2.29)
$$P_0\{T < m,\ |W(m)| < cm^{1/2}\} = (b - b^{-1})\varphi(b)\log(mc^2/m_0 b^2)$$
$$+ b^{-1}\varphi(b)[3 - (bc^{-1})^2] + o(b^{-1}\varphi(b)).$$

Miller and Siegmund (1982) discuss the history of the special case $c = b$ of (2.29), which has been given incorrectly several times in the literature.

Using methods introduced by Woodroofe (1976b, 1982), Woodroofe and Takahashi (1982) obtain the comparable approximation for $P_0\{T \leq m\}$ in the discrete case. The result is quite complicated and does not appear to yield generally more accurate approximations than the one suggested here (i.e. the sum of (2.18) with $c = b$ and $P_0\{|S_m| \geq bm^{1/2}\} = 2[1 - \Phi(b)]$).

28

3. Other Boundary Crossing Problems

In Part 3 we consider a number of somewhat related (fixed sample) statistical problems which involve boundary crossing probabilities. For historical reasons the Kolmogorov-Smirnov and Anderson-Darling statistics are discussed briefly in Section 3.1. Section 3.2 is concerned with the mathematically similar but conceptually different problem of maximum χ^2 statistics. Sections 3.3–3.6 on change point problems are the primary focus of the chapter. (These sections can be read independently of the first two.) As we shall see, the methods of Part 2 occasionally deliver an appropriate approximation immediately, sometimes additional work is required, and sometimes completely new methods are needed.

3.1 Kolmogorov-Smirnov and Anderson-Darling Statistics

Let $u_1, u_2, \cdots$ be independent and uniform on $[0,1]$, and let

$$F_n(x) = n^{-1} \sum_1^n I_{\{u_k \leq x\}}$$

be the empirical distribution function. As stated in the introduction, essentially the first boundary crossing problem in statistics is that of finding the distribution of the one-sample Kolmogorov-Smirnov statistic,

$$\sup_x [x - F_n(x)].$$

The distribution can be evaluated exactly (e.g. Birnbaum and Tingey, 1951), but the result is quite complicated. From the representation of the uniform order statistics as W_k/W_{n+1}, $k = 1, 2, \cdots, n$, where $W_k = y_1 + \cdots + y_k$ with $y_1, y_2, \cdots$ independent standard exponential, it follows that

$$P\{\sup_x [x - F_n(x)] \geq \varsigma\} = P\{\max_{1 \leq j \leq n} [W_j - j] \geq n\varsigma - 1 \mid W_{n+1} - (n+1) = -1\}.$$

The methods of Part 2 yield a large deviation and a corrected diffusion approximation, both of which are very accurate. See Siegmund (1982), Yuh (1982), and Siegmund (1984). Of course, the limiting distribution is given by (2.5) with $\xi = 0$, $m = 1$, and $b = \varsigma n^{1/2}$, but it is not a particularly good approximation for small n.

Since the Kolmogorov-Smirnov statistic is insensitive to departures in the tails from the hypothesized distribution, Anderson and Darling (1952) proposed the goodness of fit

statistic (two-sided alternative)

$$(3.1) \qquad n^{1/2} \sup_{\epsilon_1 \leq x \leq 1-\epsilon_2} \{|F_n(x) - x|/[x(1-x)]^{1/2}\} \qquad (0 < \epsilon_1 < 1 - \epsilon_2 < 1)$$

and observed that the asymptotic distribution of (3.1) as $n \to \infty$ is that of the random variable

$$(3.2) \qquad \max_{\epsilon_1 \leq t \leq 1-\epsilon_2} |W_0(t)|/[t(1-t)]^{1/2},$$

where $W_0(t)$, $0 \leq t \leq 1$, is a Brownian bridge. It is immediately verified by checking the covariance function that

$$(3.3) \qquad W(t) = (1+t)W_0[t/(1+t)] \qquad (0 \leq t < \infty)$$

is a standard (driftless) Brownian motion process, so

$$(3.4) \quad P\{\max_{\epsilon_1 \leq t \leq 1-\epsilon_2} |W_0(t)|/[t(1-t)]^{1/2} \geq b\} = P\{\max_{\epsilon_1(1-\epsilon_1)^{-1} \leq t \leq \epsilon_2^{-1}(1-\epsilon_2)} t^{-1/2}|W(t)| \geq b\}.$$

Hence the asymptotic significance level for the Anderson-Darling statistic equals the significance level of a repeated significance test for the drift of Brownian motion. In principle one can compute (3.4) exactly (e.g. DeLong, 1981), but since the answer is very complicated and is only a crude approximation to the probability of interest, a good and easily evaluated approximation seems preferable.

Since for any $r > 0$, $r^{-1/2}W(rt)$, $0 \leq t < \infty$, is again a standard Brownian motion, it follows that

$$P\{\max_{u \leq t \leq v} t^{-1/2}|W(t)| \geq b\}$$

depends only on the ratio vu^{-1}, not the actual values of u and v. Consequently by (3.4) and (2.29) with $c = b$, as $b \to \infty$

$$(3.5) \qquad \begin{aligned} &P\{|W_0(t)| \geq b[t(1-t)]^{1/2} \text{ for some } \epsilon_1 \leq t \leq 1 - \epsilon_2\} \\ &\qquad = (b - b^{-1})\varphi(b)\log[(1-\epsilon_1)(1-\epsilon_2)/\epsilon_1\epsilon_2] + 4b^{-1}\varphi(b) + c(b^{-1}\varphi(b)). \end{aligned}$$

Comparison of (3.5) with the exact numerical computations of DeLong (1981) show that it is quite accurate, even when the probability is not close to 0.

3.2 Maximum χ^2 Statistics

The random variable (3.2) arises as a limit in distribution in a context which at first appears to be quite different than the Anderson-Darling statistic.

Suppose that a 2×2 table is obtained from a categorical variable A or A^c (not A) and a dichotomized quantitative variable Y, which divides a population according to low $(Y \geq y)$ and high $(Y > y)$ values of Y. See Figure 4.

Figure 4

This situation might arise if $A(A^c)$ denotes the occurrence (non-occurrence) of some event or presence (absence) of some disease and Y is a diagnostic predictor of the event or disease. We seek a cut point y^*, which divides the population into low risk and high risk groups.

An apparently common *ad hoc* procedure for choosing y^* is obtained by the following reasoning. For a given value of y, one measure of dependence between the categories A and $Y \leq y$ is the χ^2 statistic

$$\chi_y^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)},$$

and larger values of χ_y^2 indicate a larger degree of dependence. Hence we choose y^* to maximize χ_y^2 (subject to keeping some minimal percentage of the total sample in the $Y \leq y^*$ and $Y > y^*$ categories). To assess the "significance" of this predictor, we consider the distribution of $\max_y \chi_y^2$ under the assumption of independence in the 2×2 table for all y.

Let $F_1(y) = P\{Y \leq y \mid A\}$, $F_2(y) = P\{Y \leq y \mid A^c\}$. The natural nonparametric estimators of F_1 and F_2 are

$$\hat{F}_1(y) = a/(a+b) \quad \text{and} \quad \hat{F}_2(y) = c/(c+d).$$

The hypothesis of independence in the 2×2 table for all y is $H_0 : F_1 = F_2$, and $\hat{F}(y) = (a+c)/N$ estimates the common distribution function under H_0. In terms of $\hat{F}_1$, $\hat{F}_2$, and

31

$\hat{F}$, the square root of the maximum χ^2 statistic is

$$\max_y \chi_y = \max_{c_1 \le \hat{F}(y) \le 1-c_2} |\hat{F}_1(y) - \hat{F}_2(y)|/\left\{\hat{F}(y)[1-\hat{F}(y)]\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{1/2},$$

where $n_1 = a + b$, $n_2 = c + d$. This is the natural definition of a two-sample Anderson-Darling statistic, which under H_0 converges in law to (3.2) as $\min(n_1, n_2) \to \infty$. See Miller and Siegmund (1982) for a more complete discussion and numerical examples.

Although the probabilistic aspect of this problem is already solved, natural and simple generalizations to deal with more than one predictor variable dichotomized, say, by a hyperplane seem extremely difficult. See Halpern (1982) for a more precise formulation and Monte Carlo study.

3.3 Introduction to Change Point Problems

In these final four sections we shall discuss detection and estimation of the time(s) of an abrupt change in the distribution of a sequence of observations $x_1, x_2, \cdots$. To simplify the discussion, assume that the x_i are independent and normally disributed with means $\mu^{(i)}$ and variance 1. Change point problems appear to have arisen originally in the context of quality control, where one observes the output of a production process sequentially and wants to signal any departure of the average output, from some known target value μ_0. Outstanding contributions in a long line of papers on sequential detection are Page (1954), Shiryayev (1963), Lorden (1970), and Pollak (1985, to appear).

In the following we consider only fixed sample problems involving a finite sequence $x_1, x_2, \cdots, x_m$. The specific problems to be discussed are to test the null hypothesis of no change $H_0 : \mu^{(1)} = \cdots = \mu^{(m)}$ against the alternatives of exactly one change,

$$H_1 : \exists 1 \le \rho < m \text{ such that } \mu^{(1)} = \cdots = \mu^{(\rho)} = \mu_0 \ne \mu_1 = \mu^{(\rho+1)} = \cdots = \mu^{(m)},$$

or against the "epidemic" or "square wave" alternative,

$$H_1 : \exists 1 \le \rho_1 < \rho_2 < m \text{ such that } \mu^{(1)} = \cdots = \mu^{(\rho_1)} = \mu_0,$$
$$\mu^{(\rho_1+1)} = \cdots = \mu^{(\rho_2)} = \mu_0 + \delta, \ \mu^{(\rho_2+1)} = \cdots = \mu^{(m)} = \mu_0.$$

We also consider estimation of ρ by a confidence set when the hypothesis of exactly one

32

change is assumed to be true. Typically μ_0 and $\delta = \mu_1 - \mu_0$ are unknown, but it sometimes seems reasonable to suppose that a particular value of δ is a minimum threshold of interest and hence to regard δ as known for the purpose of deriving a test statistic.

Examples of change point problems in epidemiology are described by Worsley (1983) and by Levin and Kline (1984). Here one is interested in testing whether the incidence of a disease has remained constant over time, and if not, in estimating the time(s) of change(s) in order to suggest possible causes. Kendall and Kendall (1980) describe an interesting change point problem in archaeology, and Brown, Durbin, and Evans (1975) give a number of econometric examples.

In Section 3.4 we consider the likelihood ratio test of no change against the alternative of exactly one change. A large number of test statistics have been proposed for this problem, and there is no attempt to compare them here. The main conclusion is that the methods of Part 2 provide the basic tools to study a number of these tests without resorting to the numerical or Monte Carlo efforts that have been the basis of earlier studies (e.g. Sen and Srivastava, 1975).

Section 3.5 is concerned with finding a confidence set for ρ. In the case where μ_0 and μ_1 are both known, we compare confidence intervals based on the maximum likelihood estimator $\hat{\rho}$ and confidence sets (which generally are not intervals) based directly on the likelihood function. Hinkley (1970, 1972) has mentioned both methods, but he directs his efforts primarily at computational problems and does not compare the two methods quantitatively. To minimize the computational difficulties and facilitate a simple comparison, we consider the case of Brownian motion. The likelihood based method appears to be preferable, and it is extended to the case of unknown nuisance parameters, μ_0 and μ_1.

Section 3.6 is concerned with testing the hypothesis of no change against an epidemic alternative. Here one encounters processes with a multidimensional indexing set, which introduce some new problems. The methods of Part 2 can be used in some special cases, but in others an adaptation of ideas of Bickel and Rosenblatt (1973) or Qualls and Watanabe (1973) seems more fruitful.

3.4 Tests Against the Alternative of Exactly One Change

The problem of testing the null hypothesis of no change, $H_0 : \mu^{(1)} = \cdots = \mu^{(m)} = \mu_0$, against the alternative of exactly one change, $H_1 : \exists 1 \le \rho < m$ such that $\mu^{(1)} = \cdots = \mu^{(\rho)} = \mu_0 \ne \mu_1 = \mu^{(\rho+1)} = \cdots \mu^{(m)}$ (μ_0 and μ_1 both unknown) has been widely discussed; and a number of test statistics have been proposed. The quasi-Bayesian statistics of Chernoff and Zacks (1964) and Gardner (1969) are analytically tractable, but maximum likelihood type statistics have typically been studied by numerical or Monte Carlo methods (e.g. Sen and Srivastava, 1975, Worsley, 1983).

The square root of the log likelihood ratio statistic is proportional to

$$(3.6) \qquad \max_{1 \le k < m} \{|S_k - k S_m/m|/[k(1 - k/m)]^{1/2}\}.$$

A simple heuristic derivation of this statistic with minimal calculation is to suppose momentarily that H_1 specifies $\rho = k$. The problem then becomes a two population test to decide whether the mean (μ_0) of the first k observations equals the mean (μ_1) of the last $m - k$. The standard test statistic is the normalized difference between the mean of the first k observations, S_k/k, and the overall mean, S_m/m. This is just (3.6) without the max, which accounts for the fact that ρ is actually unknown.

Slightly more generally, we shall consider

$$(3.7) \qquad \max_{m_0 \le k \le m_1} \{|S_k - k S_m/m|/[k(1 - k/m)]^{1/2}\},$$

where $1 \le m_0 < m_1 < m$. (A justification is given below.)

To obtain some intuition for the virtues and defects of (3.7) consider also the *ad hoc* suggestion of Pettit (1980)

$$(3.8) \qquad \max_{1 \le k < m} |S_k - k S_m/m|.$$

Under H_0 the process $S_k - k S_m/m$, $k = 0, 1, \cdots m$ is the same as the conditonal process S_k, $k = 0, 1, \cdots, m$ given that $S_m = 0$, i.e. the same as a Brownian bridge observed at discrete instants of time. Hence an excellent approximation to the significance level of (3.8) can be obtained from (2.11) or (2.16) (multiplied by 2 to account for the two-sided alternative); the significance level of (3.7) is discussed below.

Under H_1 the drift of $S_k - k\,S_m/m$, $k = 0, 1, \cdots, m$, is

$$k(1 - \rho/m)(\mu_1 - \mu_0), \qquad k \leq \rho$$

(3.9)

$$(\rho/m)(m - k)(\mu_1 - \mu_0), \qquad k \geq \rho,$$

and the residual process after subtracting out the drift is again a Brownian bridge observed at discrete instants of time.

It seems intuitively clear from (3.9) as illustrated in Figure 5 that (3.8) is more powerful than (3.6) for detecting changes that occur near $m/2$, whereas the converse is true for changes occurring near the endpoints 0 and m.

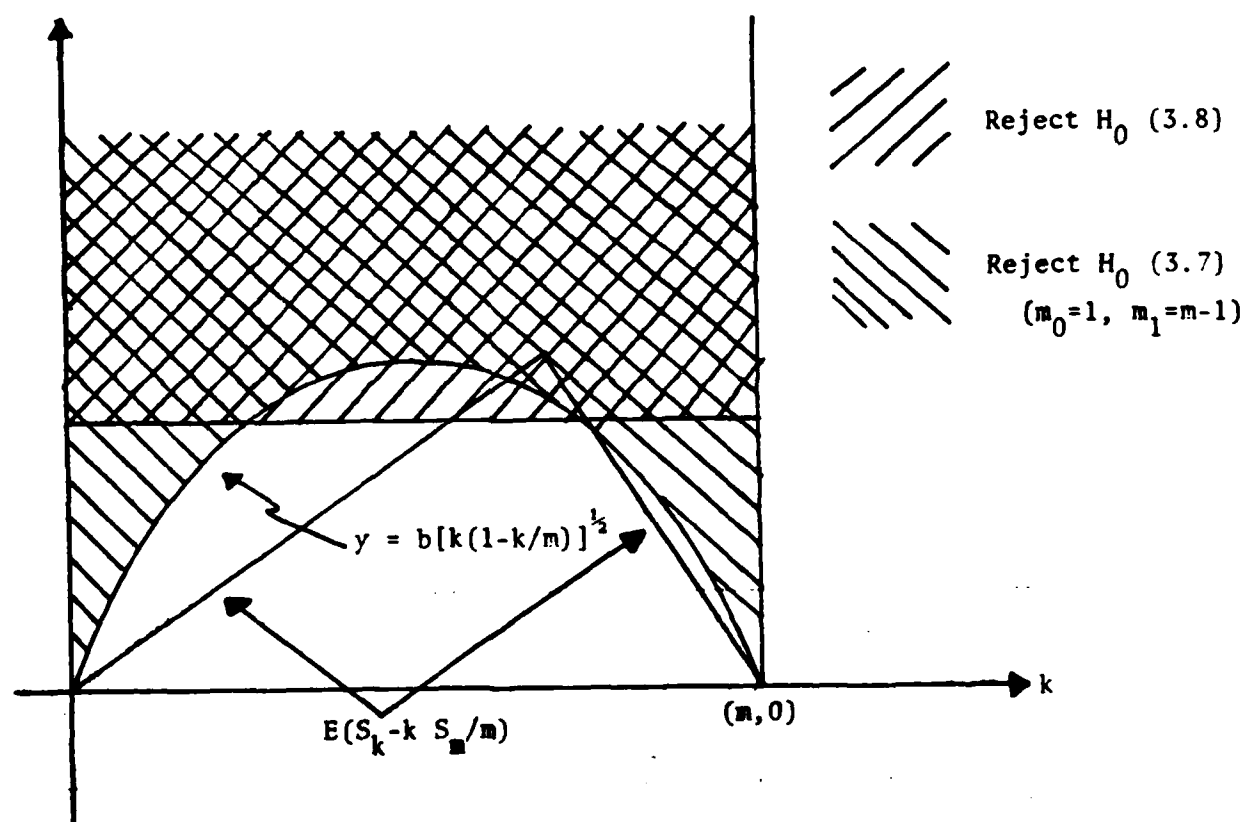**Figure 5**

It is intrinsically difficult to detect a change that occurs near one or the other endpoint, and the likelihood ratio statistic pays for its efforts to do so by giving up power near $\rho = m/2$. The introduction of m_0 and m_1 in (3.7) gives the statistician the flexibility to give up some power to detect changes occurring near the endpoints in return for an increase in power

35

near $m/2$.

By conditioning on S_ρ, one can obtain approximations to the power of (3.7) and (3.8), which can be used to compare these statistics with each other and with other proposals (e.g. the recursive residual test of Brown, Durbin, and Evans, 1975). A more complete discussion will appear in a future publication of B. and K. James. To illustrate the applicability of the methods of Part 2, and to prepare for the discussion of confidence sets in Section 3.5, an approximation to the significance level of (3.7) is given below.

Let $z_1, z_2, \cdots, z_m$ be independent standard normal random variables, and put $S_n = z_1 + \cdots + z_n$. We continue to use the notation

$$P_\xi^{(n)}(A) = P(A \mid S_n = \xi), \qquad A \in \mathcal{J}(z_1, \cdots, z_n).$$

Let $b > 0$, $m = 2, 3, \cdots$, $1 \le m_0 < m$, and define

$$(3.9) \qquad T = \inf\{n : n \ge m_0, \ |S_n| \ge b[n(1 - n/m)]^{1/2}\}.$$

Let $m_0 \le m_1 \le m - 1$. The significance level of the test defined by (3.7) is

$$(3.10) \qquad \begin{aligned} P_0^{(m)}\{T \le m_1\} &= P_0^{(m)}\{|S_{m_1}| \ge b[m_1(1 - m_1/m)]^{1/2}\} \\ &+ \int_{|\xi| < b[m_1(1 - m_1/m)]^{1/2}} P_\xi^{(m_1)}\{T < m_1\} P_0^{(m)}\{S_{m_1} \in d\xi\}. \end{aligned}$$

Theorem 3.11. Assume that $b \to \infty$, $m_0 \to \infty$, $m_1 \to \infty$, $m \to \infty$ in such a way that for some $0 \le t_0 < t_1 < 1$ and $\mu_1 > 0$

$$m_i/m \to t_i \ (i = 0, 1) \quad \text{and} \quad b/m^{1/2} = \mu_1.$$

Let $\xi = m\xi_0$ for some $|\xi_0| \in (\mu_1(1 - t_1)[t_0/(1 - t_0)]^{1/2}, \ \mu_1[t_1(1 - t_1)]^{1/2})$.

Then as $m \to \infty$, $P_\xi^{(m_1)}\{T < m_1\} \sim$

$[t_1(1 - t_1)]^{1/2} \mu_1 \xi_0^{-1} \nu[\mu_1^2(1 - t_1)/\xi_0 + \xi_0/(1 - t_1)] \exp\{-\frac{1}{2}m[\mu_1^2 - \xi_0^2/t_1(1 - t_1)]\}$, where ν is given by (2.12).

Remarks. Substitution of this asymptotic expression into (3.10) suggests the approximation

$$(3.12) \qquad P_0^{(m)}\{T \le m_1\} \cong 2b \, \varphi(b) \int_{b(m_1^{-1} - m^{-1})^{1/2}}^{b(m_0^{-1} - m^{-1})^{1/2}} x^{-1}\nu(x + b^2/mx)dx + 2[1 - \Phi(b)],$$

36

which can be shown to be a valid asymptotic relation (even if m_0 and $m - m_1$ are $o(m)$). A proof of Theorem 3.11 along the lines of Theorem 2.18 has an interesting twist, leading to some new technical problems. An informal discussion is contained in Appendix 1. Siegmund (1985, Chapter XI) derives (3.12) directly without first obtaining Theorem 3.11. However, we shall find Theorem 3.11 to be of interest in its own right in the next section.

Table 5 gives an indication of the accuracy of (3.12). For comparison an exact numerical calculation from Worsley (1983) or the result of a Monte Carlo experiment plus or minus one standard error is also given. There were 2500 repetitions of the Monte Carlo experiment, and importance sampling along the lines discussed in Siegmund (1975) was used for variance reduction.

Table 5

$$P_0^{(m)}\{T \leq m_1\}$$

b	m_0	m_1	m	Probability Approximation (3.12)	Exact or Monte Carlo
2.91	1	3	4	.010	.01
2.65	1	9	10	.052	.05
2.38	1	9	10	.105	.10
2.38	4	8	10	.057	.058 ± .002
2.5	1	19	20	.120	.110 ± .002
2.5	1	10	20	.066	.063 ± .001
2.5	4	16	20	.073	.074 ± .002

Exact values from Worsley (1983)

3.5 Confidence Sets for ρ

This section is concerned with finding confidence sets for ρ, when μ_0 and μ_1 are regarded as nuisance parameters. Initially we shall asume that μ_0 and μ_1 are both known, a case studied in considerable detail by Hinkley (1970, 1972), who suggested a method based on the maximum likelihood estimator $\hat{\rho}$ and a second method based directly on the likelihood function. In order to simplify the computational difficulties as much as possible and obtain a picture of the relative merits of these two proposals, we begin with the case of

Brownian motion observed for $0 \leq t \leq m$. As the results of Part 2 indicate, use of Brownian motion as an approximation usually yields quantitatively poor results for boundary crossing probabilities. For comparing competing procedures, however, Brownian motion can be quite useful.

Hence let $W(t)$, $0 \leq t \leq m$, be standard Brownian motion, and assume that the observed process $X(t)$, $0 \leq t < m$, satisfies

$$dX(t) = \mu_0 \, dt + dW(t) \quad \text{for} \quad 0 < t < \rho$$
$$= \mu_1 \, dt + dW(t) \quad \text{for} \quad \rho < t < m,$$

where μ_0 and μ_1 are both known and ρ is unknown. There is no loss of generality in taking $\mu_0 = 0$. Put $\delta = \mu_1$. The likelihood function at $\rho = t$ is proportional to $\exp\left[\frac{1}{2}\delta^2 t - \delta X(t)\right]$. Hence the log likelihood $\ell(t) = \frac{1}{2}\delta^2 t - \delta X(t)$ satisfies

$$d\ell(t) = \frac{1}{2}\delta^2 dt - \delta \, dW(t), \qquad 0 < t < \rho$$
$$= -\frac{1}{2}\delta^2 dt - \delta \, dW(t), \quad \rho < t < m,$$

i.e. $\ell(t)$ is Brownian motion with drift $\frac{1}{2}\delta^2$ or $-\frac{1}{2}\delta^2$ according t is $< \rho$ or $> \rho$, and $\hat{\rho}$ is the time at which this process takes on its maximum value.

It is easy to compute the distribution of $\hat{\rho}$, but to simplify the resulting expression we assume that ρ and $m - \rho$ are effectively infinitely large. Consider

$$P_\rho\{\hat{\rho} - \rho \in (t, \ t + dt), \ \ell(\hat{\rho}) - \ell(\rho) \in (z, \ z + dz)\}.$$

This joint density can be evaluated by (i) conditioning on $\ell(t) - \ell(\rho) = z - y(dt)^{1/2}$, $\ell(t + dt) - \ell(\rho) = z - z(dt)^{1/2}$, (ii) computing the (conditional) probability that the process $\ell(s) - \ell(\rho)$ does not attain the value z for $s < t$ nor for $s > t + dt$ and its maximum in the interval $(t, \ t + dt)$ is in $(z, \ z + dz)$, and (iii) integrating out y and z over $(0, \infty)$. See Figure 6. The joint density is $\delta^{-1} z |t|^{-3/2}[1 - \exp(-z)]\varphi\left(z/\delta|t|^{1/2} + \frac{1}{2}\delta|t|^{1/2}\right) dz \, dt$ for $z > 0$ and 0 otherwise. Integration over $z \in (0, \infty)$ and $t \in (r, \infty)$ $(r > 0)$ yields

$$
(3.13) \qquad
\begin{aligned}
P_\rho\{\hat{\rho} - \rho > r\} &= \Phi\left(-\frac{1}{2}\delta r^{1/2}\right)\left(\frac{1}{2}\delta^2 r + \frac{5}{2}\right) - \delta r^{1/2}\varphi\left(\frac{1}{2}\delta r^{1/2}\right) \\
&\quad - \left(\frac{3}{2}\right)\exp(\delta^2 r)\Phi(-3\delta r^{1/2}/2).
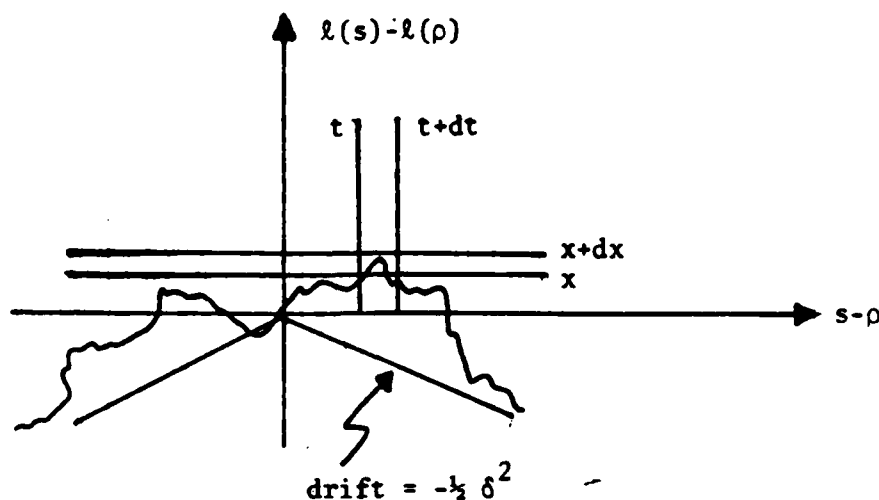\end{aligned}
$$

Figure 6

It follows from (3.13) that the length of a 95% confidence interval for ρ obtained by treating $\hat{\rho} - \rho$ as a pivotal quantity is about $22/\delta^2$. (Without the assumption that ρ and $m - \rho$ are infinitely large, $\hat{\rho} - \rho$ would not be an exact pivotal).

The definition of a likelihood based confidence set is very simple. For $z > 0$ let $A(\rho) = \{\sup_s[\ell(s) - \ell(\rho)] < z\}$. Choose z so that $P_\rho\{A(\rho)\} = 1 - \alpha$ and define the confidence set to be those values t for which $\ell(t) > \ell(\hat{\rho}) - z$. Again assuming that ρ and $m - \rho$ are effectively infinite, one easily sees that (cf. Figure 6)

$$P_\rho\{A(\rho)\} = [1 - \exp(-z)]^2,$$

so

$$z = -\log[1 - (1 - \alpha)^{1/2}].$$

Observe that the confidence set obtained in this manner is by no means an interval. In fact, because of the rapid fluctuations of Brownian sample paths, with probability one it consists of the union of infinitely many open intervals.

To compare this likelihood based confidence set with the confidence interval determined above, we compute the expected size of the confidence set, i.e.

$$(3.14) \qquad E_\rho[\lambda\{t : \omega \in A(t)\}] = \int_{-\infty}^{\infty} P_\rho[A(t)]dt,$$

where λ denotes Lebesgue measure and ω is a sample path $\ell(s) - \ell(\rho)$, $-\infty < s < \infty$. By conditioning on $\ell(t)$ one can derive an expression for $P_\rho[A(t)]$, which when integrated shows that (3.14) equals

$$4\delta^{-1}[1 - \exp(-z)]\{z\delta^{-1} - (2\delta)^{-1}[1 - \exp(-z)]\}$$
$$= 4\delta^{-2}(1 - \alpha)^{1/2}\{-\log[1 - (1 - \alpha)^{1/2}] - (1 - \alpha)^{1/2}\}.$$

For a 95% confidence set the expected size is about $10.5/\delta^2$, or less than one half the length of the corresponding confidence interval based on the distribution of $\hat{\rho}$. Numerical evidence indicates that the likelihood based set has approximately the same expected size advantge throughout the range of commonly used confidence levels.

Remark. Cobb (1978) has proposed yet a third confidence set (interval) for ρ, which is in a certain respect intermediate between the two proposed here. Given a suitable $t_0 > 0$, Cobb treats $\ell(\hat{\rho}) - \ell(t)$, $|t - \hat{\rho}| \le t_0$, as ancillary and bases his interval on the conditional distribution of $\hat{\rho} - \rho$ given this ancillary statistic. Thus, if the likelihood function drops off sharply from its maximum at $\hat{\rho}$, Cobb's interval is short – a property shared by the likelihood based confidence set. There is some arbitrariness in the choice of t_0, which seems a definite disadvantage if one tries to adapt this method to the case of unknown μ_0 and μ_1, especially if m is of moderate size. Nonetheless, it would be interesting to compare Cobb's method with those described above.

Now suppose that our observations are $z_1, \cdots, z_m$ as in Section 3.4, that the hypothesis of exactly one change is true, and that μ_0 and μ_1 are unknown nuisance parameters. A likelihood based confidence set for ρ can be defined as follows. The log likelihood ratio statistic for testing the hypothesis $\rho = \rho_0$ against the alternative of arbitrary ρ is (cf. (3.6))

$$\max_k[(S_k - k\,S_m/m)^2/k(l - k/m)] - (S_{\rho_0} - \rho_0\,S_m/m)^2/\rho_0(1 - \rho_0/m).$$

Hence for $1 \leq m_0 < m_1 < m$ and $c > 0$ define the events

$$(3.15) \quad A(\rho, c) = \{ \max_{m_0 \leq k \leq m_1} [(S_k - k\,S_m/m)^2/k(1-k/m)] - (S_\rho - \rho\,S_m/m)^2/\rho(1-\rho/m) \leq c^2 \}.$$

Although the unconditional probability of $A(\rho, c)$ depends on both ρ and $\delta = \mu_1 - \mu_0$, its conditional probability given that $S_\rho - \rho\,S_m/m = \xi$, say, does not depend on δ. (Conditionally, $S_k - k\,S_m/m - \xi k/\rho$, $k = 0, 1, \cdots, \rho$ and $S_k - k\,S_m/m - \xi(m-k)/(m\rho)$, $k = \rho, \rho+1, \cdots, m$ are two stochastically independent Brownian bridges in discrete time.) Hence one can in principle determine $c = c(\alpha, \rho, \xi)$ such that

$$(3.16) \quad P_\rho\{A(\rho, c) \mid S_\rho - \rho S_m/m = \xi\} = 1 - \alpha.$$

From (3.16) it follows immediately that the set of all ρ such that the sample path $\omega = \{S_k - k\,S_m/m,\ k = 0, 1, \cdots, m\}$ belongs to

$$A[\rho, c(\alpha, \rho, S_\rho - \rho\,S_m/m)]$$

is a $(1 - \alpha)100\%$ confidence set for ρ.

To implement this procedure one must compute the conditional probability in (3.16); but this problem is already solved (asymptotically) in Theorem 3.11, as follows.

In terms of the stopping time T defined in (3.9)

$$(3.17) \quad \begin{aligned} P_\rho\{A(\rho, c) \mid S_\rho - \rho S_m/m = \xi\} &= P_\xi^{(\rho)}\{T < \rho\} + P_\xi^{(m-\rho)}\{T < m - \rho\} \\ &\quad - [P_\xi^{(\rho)}\{T < \rho\} P_\xi^{(m-\rho)}\{T < m - \rho\}], \end{aligned}$$

where $b = [c^2 + \xi^2/\rho(1-\rho/m)]^{1/2}$. If in addition to the asymptotic normalization of Theorem 3.11 one assumes that c^2 is proportional to m and ρ/m equals some constant in $(0, 1)$, then Theorem 3.11 and (3.17) yield

$$(3.18) \quad \begin{aligned} P_\rho\{A(\rho, c) \mid S_\rho - \rho\,S_m/m = \xi\} &\sim \exp\left(-\frac{1}{2}c^2\right)[1 + \rho(1 - \rho/m)c^2/\xi^2]^{1/2} \\ &\quad \cdot \{\nu[c^2(1 - \rho/m)/\xi + \xi/\rho(1 - \rho/m)] + \nu[c^2\rho/m\xi + \xi/\rho(1 - \rho/m)]\}, \end{aligned}$$

where ν is defined in (2.12) and evaluated approximately in (2.13)–(2.14).

Table 6 indicates the accuracy of (3.18). To obtain a Monte Carlo estimate of the desired probability, importance sampling (Siegmund, 1975) was used to obtain independent

estimates of the two probabilities on the right hand side of (3.17). The standard error of the overall estimate was obtained via the obvious Taylor series expansion. The number of repetitions in each Monte Carlo experiment was 900.

<div align="center">

Table 6

$P_\rho\{A(\rho,c) \mid S_\rho - \rho \, S_m/m = \xi\}$

</div>

m	c	ρ	$\xi_0 = \xi m^{-1}$	Probability Approximation (3.18)	Monte Carlo Estimate
40	2.5	20	.50	.027	.029 ± .001
40	2.4	12	.25	.059	.058 ± .001
20	2.4	10	.25	.066	.059 ± .001
20	2.4	6	.25	.057	.052 ± .001
20	2.2	6	.50	.043	.042 ± .002

3.6. Tests Against the Epidemic Alternative

In this section we consider several tests of the hypothesis of no change $H_0 : \mu^{(1)} = \mu^{(2)} = \cdots = \mu^{(m)} = \mu_0$, against the epidemic or square wave alternative, $H_1 : \exists \, 1 \leq \rho_1 < \rho_2 < m$ such that $\mu^{(1)} = \cdots = \mu^{(\rho_1)} = \mu_0$, $\mu^{(\rho_1+1)} = \cdots = \mu^{(\rho_2)} = \mu_0 + \delta$, $\mu^{(\rho_2+1)} = \cdots = \mu^{(m)} = \mu_0$. Results for this problem are incomplete, and our goals will be (i) to show that these tests naturally involve new boundary crossing problems and to (ii) suggest possible approaches to their solution. The problems are different than those discussed earlier in this paper, and the methods of Part 2 seem of limited usefulness. A promising alternative approach is provided by the method of Pickands (1969) as developed independently by Bickel and Rosenblatt (1973) and Qualls and Watanabe (1973). The results presented here are joint work with M. Hogan, which will be described in greater detail in a future publication.

Typically μ_0 and δ are unknown nuisance parameters, although often only one-sided alternatives with $\delta > 0$, say, are of interest. We shall assume that there is some threshold change, δ_0, which one is interested in detecting and consider tests for the particular alternative $\delta = \delta_0$. Thus in effect we assume that δ is known for the purpose of deriving a test statistic, although a complete evaluation of that statistic would involve all values of δ, not

just the hypothesized value δ_0.

In contrast to the alternative of exactly one change, the epidemic alternative has rarely been considered. See Levin and Kline (1984) and Bhattacharya and Brockwell (1976) for two quite different discussions.

Assume that $\delta = \delta_0$ is known. In the unlikely case that μ_0 is also known the log likelihood ratio statistic for testing H_0 against H_1 is proportional to

$$(3.19) \quad \begin{aligned} Z_1 &= \max_{0 \le i < j \le m} \left[S_j - j\,\mu_0 - (S_i - i\mu_0) - \frac{1}{2}\delta_0(j - i) \right] \\ &= \max_{1 \le j \le m} \left[S_j - j\,\mu_0 - j\,\delta_0/2 - \min_{0 \le i < j}(S_i - i\mu_0 - i\delta_0/2) \right]. \end{aligned}$$

For the case of unknown μ_0 Levin and Kline (1984) suggest the use of (3.19) with μ_0 replaced by its maximum likelihood estimate under H_0, namely $\hat{\mu}_0 = m^{-1}S_m$, to obtain

$$(3.20) \quad Z_2 = \max_{0 \le i < j \le m} [S_j - j\,S_m/m - (S_i - iS_m/m) - (j - i)\delta_0/2].$$

The actual log likelihood ratio statistic in the case of known $\delta = \delta_0$ and unknown μ_0 is easily calculated to be

$$(3.21) \quad Z_3 = \max_{0 \le i < j \le m} \left\{ S_j - S_i - (j - i)S_m/m - \frac{1}{2}\delta_0(j - i)[1 - (j - i)/m] \right\}.$$

Levin and Kline (1984) discuss Bernoulli and Poisson data; and in that context an important aspect of their test is their proposal to use a conditional distribution given S_m, which under H_0 does not depend on the unknown μ_0, to compute a significance level. In the Gaussian case under discussion here the conditional and unconditional distributions are the same. Since (3.20) is somewhat easier to study than (3.21), an interesting question is to what extent the two statistics behave similarly. Presumably they do if the duration of the epidemic, $\rho_2 - \rho_1$, is small compared to m, but not in general.

For a completely different problem which leads to consideration of (3.20) in the special case $\delta_0 = 0$ and the simpler framework of continuous time, see Adler and Brown (1984). For a different underlying random walk the probability that Z_1 in (3.19) is greater than b can be interpreted as the probability that among the first m customers of a $G/GI/1$ queue, at least one has a waiting time exceeding b.

The appearance of a two-dimensional indexing set in (3.19)-(3.21), corresponding to the unknown onset and disappearance of the epidemic, makes the null hypothesis sampling distributions of these statistics quite different from those discussed earlier. Approximations to the power function seem more complicated in detail, but for the most interesting range of parameter values do not seem to require fundamentally new ideas. The remainder of this section describes some promising methods for approximating the null hypothesis distributions of (3.19)-(3.21).

We begin with the relatively simple (3.19), which gives us an idea of what $\overline{\text{we}}$ can hope to achieve in the more complicated (3.20) and (3.21).

Let $y_1, y_2, \cdots$ be independent, identically distributed random variables with $E(y_i) < 0$. Let $\tilde{S}_n = y_1 + \cdots + y_n$ and for $b > 0$ define

$$(3.22) \qquad \tau = \tau(b) = \inf\{n : \tilde{S}_n - \min_{0 \leq k \leq n} \tilde{S}_k \geq b\}.$$

The following inequality is useful in analyzing (3.19).

Proposition 3.23. Let $\tau = \tau(b)$ be defined by (3.22), $\tau_+ = \inf\{n : \tilde{S}_n > 0\}$, and $T = \inf\{n : \tilde{S}_n \notin (0, b]\}$. Then

$$P\{\tau(b) \leq m\} \leq P\{\tau_+ = \infty\}E\{(m - T + 1); T < m, S_T \geq b\}$$
$$(3.24) \qquad\qquad + \sum_{n=0}^{m-1} P\{n < \tau_+ < \infty\}P\{T < m - n, S_T \geq b\}.$$

Moreover, a lower bound for $P\{\tau(b) \leq m\}$ is the right hand side of (3.24) divided by $1 + E\{(m - T + 1); T < m, S_T \geq b\}$.

Remark 3.25. With large deviation scaling and observations whose distribution can be imbedded in an exponential family, one can use likelihood ratio identities similar to, but simpler than, those developed in Part 2 to obtain first or second order asymptotic approximations to $P\{\tau(b) \leq m\}$. For example, for the normal case in (3.19), if $m \exp(-\delta_0 b) \to 0$ and $\delta_0 m / 2b - 1$ is bounded below by some positive number, then

$$(3.26) \qquad P_{\mu_0}\{Z_1 \geq b\} \sim \delta_0(m\delta_0/2 - b)\nu^2(\delta_0)\exp(-\delta_0 b),$$

where ν is given by (2.12). If in fact $m^2 \exp(-\delta_0 b) \to 0$, then the error in (3.26) is $K(\delta_0)\exp(-\delta_0 b)(1 + o(1))$, where K is very complicated to evaluate exactly, but satis-

fies $K(\delta) \sim 2^{1/2}\delta/8$ as $\delta \to 0$ (and better approximations are possible). Details of this asymptotic analysis will be presented elsewhere.

Proof of Proposition 3.23. Let ω_n^+ denote the n-shifted sample path, so $\tilde{S}_k(\omega_n^+) = \tilde{S}_{n+k}(\omega) - \tilde{S}_n(\omega)$. The event $\{\tau \leq m\}$ can be decomposed into a union of disjoint events, $\{\tau \leq m\} = \bigcup_{n=0}^{m-1}\{\tau > n, \ \tilde{S}_n = \min_{0 \leq k \leq n}\tilde{S}_n, \ T(\omega_n^+) \leq m - n, \ \tilde{S}_T(\omega_n^+) \geq b\}$. Hence by independence,

$$P\{\tau \leq m\} = \sum_{n=0}^{m-1} P\{\tau > n, \ \tilde{S}_n = \min_{k \leq n}\tilde{S}_k\}P\{T \leq m - n, \ \tilde{S}_T \geq b\}$$

$$\geq \sum_{n=0}^{m-1}[P\{\tilde{S}_n - \tilde{S}_k \leq 0 \quad \forall k \leq n\} - P\{\tau \leq n\}]P\{T \leq m - n, \ \tilde{S}_T \geq b\}$$

$$\geq \sum_{n=0}^{m-1}[P\{\tau_+ > n\} - P\{\tau \leq m\}]P\{T \leq m - n, \ \tilde{S}_T \geq b\}$$

$$= [P\{\tau_+ = \infty\} - P\{\tau \leq m\}]E\{(m - T + 1); \ T < m, \ \tilde{S}_T \geq b\}$$

$$+ \sum_{n=0}^{m-1} P\{n < \tau_+ < \infty\}P\{T \leq m - n, \ \tilde{S}_T \geq b\}.$$

Rearranging gives the lower bound, and a similar, simpler argument gives the upper bound.

In principle the method sketched in Proposition 3.23 and Remark 3.25 to approxiamte the distribution of (3.19) should also be applicable to (3.20). In this case because of non-stationarity one must decompose $\{Z_2 \geq b\}$ not only according to the location of a (relative) minimum of the sample path but also according to the value of the process at the minimum. The details become much more complicated and are not pursued here. For the simpler case of Brownian motion it is straightforward to obtain what one expects to be very goo. .pper bounds. Analyzing these asymptotically leads to the following conjecture.

Conjecture. Suppose $m \to \infty$, $b \to \infty$ such that for some fixed $0 < \varsigma < \infty$ and $-\infty < \xi_0 < \varsigma$,

$$b/m = \varsigma \quad \text{and} \quad \xi/m = \xi_0.$$

Then

(3.27)

$$P_\xi^{(m)}\{\max_{0 < s < t < m}[W(t) - W(s)] \geq b\} = [2m^{-1}(2b - \xi)(b - \xi) + 1 + o(1)]\exp[-2m^{-1}b(b - \xi)].$$

45

One can give a rigorous proof of the leading term in (3.27) (cf. Theorem 3.28 below), but the second term causes some difficulty. Since a standard reflection argument yields an exact evaluation of the two-sided probability,

$$P_\varsigma^{(m)}\left\{\max_{0<s<t<m}|W(t)-W(s)|\geq b\right\},$$

it is surprising that the one-sided problem should appear to be considerably more difficult.

An alternative method for approximating the null hypothesis distributions of (3.19) and (3.20), which works equally well for (3.21), is that developed independently by Bickel and Rosenblatt (1973) and Qualls and Watanabe (1973). (Both of these papers generalize to multidimensional time parameters the method of Pickands (1969) for a linear time parameter.) Since the general results of these authors give a tail probability for the maximum of a Gaussian field in terms of an integral involving another complicated probability, it is not immediately evident that the computational problem has been essentially simplified. But for Gaussian fields built up from random walks (or Brownian motion) in sufficiently simple ways one can use renewal theory to evalute the required intergrals in terms of the function ν of (2.12). For illustration the tail behavior under H_0 of (3.20) and (3.21) is given below.

Let $x_1, x_2, \cdots$ be independent standard normal random variables, and put $S_n = x_1 + \cdots + x_n$. Suppose b and $m \to \infty$ in such a way that $m^{-1}b = \varsigma$ is a fixed positive constant.

Theorem 3.28. For $\xi_0 > -\varsigma$

$$P\{\max_{0\leq i<j\leq m}[S_j - S_i - m^{-1}(j-i)S_m - (j-i)\xi_0] \geq b\}$$
$$\sim \nu^2[2(2\varsigma+\xi_0)][2m(2\varsigma+\xi_0)(\varsigma+\xi_0)]\exp[-2m\varsigma(\varsigma+\xi_0)],$$

where ν is given by (2.12).

Theorem 3.29. Let $\xi_0 \geq 0$. Then for $\xi_0 > 4\varsigma$

$$P\{\max_{0\leq i<j\leq m}[S_j - S_i - m^{-1}(j-i)S_m - \xi_0(j-i)(1-m^{-1}(j-i))] \geq b\}$$
$$\sim \nu^2(2\xi_0)m\xi_0^2\left[\frac{1}{4}-\varsigma/\xi_0\right]^{-1/2}\exp(-2m\xi_0\varsigma),$$

while for $0 \leq \xi_0 < 4\varsigma$ it is

$$\sim \nu^2(\xi_0+4\varsigma)4m(\varsigma+\xi_0/4)^{5/2}(\varsigma-\xi_0/4)^{-1/2}\exp[-2m(\varsigma+\xi_0/4)^2],$$

46

where ν is given by (2.12).

Note that in (3.26), Theorem 3.28, and Theorem 3.29 the function $\nu(\cdot)$ which accounts for excess over the boundary is squared, basically because of the two dimensional time parameter. Since typical values for $\nu(\cdot)$ are in the range .5 to .7, for these problems use of a simple Brownian motion approximation, which replaces $\nu(\cdot)$ by 1, probably gives extremely poor results.

The preceding discussion is only a beginning attempt to study the problem of change-points with epidemic alternative. It is included here to show how quickly natural generalizations of previous work lead into new territory, requiring new ideas. Two obvious questions are (i) how good are these approximations and (ii) what do they (presumably in conjunction with approximations for the power function) tell us about the relative merits of (3.20) and (3.21)? Preliminary Monte Carlo experiments indicate that the approximations are not nearly so accurate for small m as those given in Part 2, although the derivations of (3.26) and (3.27) might lead one to expect quite good approximations.

Appendix 1
Proof of Theorem 3.11

Both Theorems 2.18 and 3.11 can be proved by the method of Siegmund (1982), which requires rather length analytic calculations. The somewhat different method used in this paper to prove Theorem 2.18 yields a considerably simpler proof of that result, so one naturally asks how well it adapts to related problems. As we shall see below, it gives the appearance of a relatively computation free proof of Theorem 3.11. However, there are some technical problems which seem to demand additional analytic computation for their complete solution.

Let T be defined by (3.9) and assume the conditions of Theorem 3.11. We also use the notation $P_{\lambda,\xi}^{(m)}$ from the proof of Theorem 2.18. Let

$$(A.1) \qquad \tilde{P}_{\xi}^{(m_1)} = \int_{-\infty}^{\infty} P_{\lambda,\xi}^{(m_1)} \varphi\{[(1-t_1)/mt_1]^{1/2}[\lambda - \xi/(1-t_1)]\}[(1-t_1)/mt_1]^{1/2} d\lambda.$$

Also let $T^* = \sup\{n : n < m_1, |S_n| \geq b[n(1-n/m)]^{1/2}\}$, so $P_{0,\xi}^{(m_1)}\{T < m_1\} = P_{0,\xi}^{(m_1)}\{T^* \geq m_0\}$. As in the proof of Theorem 2.18 one easily calculates the likelihood ratio of $x_n, \cdots, x_{m_1}$ under $\tilde{P}_{\xi}^{(m_1)}$ relative to $P_{0,\xi}^{(m_1)}$ and obtains

$$[n(1-t_1)/(m-n)t_1]^{1/2} \exp\left[\frac{1}{2}S_n^2/n(1-n/m) - \frac{1}{2}\xi^2/m_1(1-m_1/m)\right],$$

where $t_i = m_i/m$ $(i = 0,1)$. Hence Wald's likelihood ratio identity yields

$$(A.2) \qquad \begin{aligned} P_{0,\xi}^{(m_1)}\{T^* \geq m_0\}[(1-t_1)/t_1]^{1/2} \exp\left[\frac{1}{2}(b^2 - m\,\xi_0^2/t_1(1-t_1))\right] \\ = \tilde{E}_{\xi}^{(m_1)}\{[(m-T^*)/T^*]^{1/2} \exp(-R_m^*); T^* \geq m_0\}, \end{aligned}$$

where $R_m^* = \frac{1}{2}[S_{T^*}^2/T^*(1 - T^*/m) - b^2]$.

The equation (A.2) is analogous to (2.23) in the proof of Theorem 2.18, and we try to evaluate it similarly. Let

$$r = \inf\{n : (\xi + S_n)^2/(m_1 - n)[1 - m^{-1}(m_1 - n)] \geq b^2\}$$

and

$$R_m = \frac{1}{2}\{(\xi + S_r)^2/(m_1 - r)[1 - m^{-1}(m_1 - r)] - b^2\}.$$

48

Past experience with large deviation scaling leads one to expect that $r/m \to t$ in probability for some constant t. We consider the definition of r expressed in the general form

$$(A.3) \qquad\qquad \tau = \inf\{n : mh(n/m,\ S_n/m) \geq 0\}$$

and expand h in a Taylor series about $(t, \mu t)$, where μ denotes the drift per unit time of S_n (from (A.1) $\mu = \xi_0/(1 - t_1)$). This shows that for n close to the random time τ

$$mh(n/m,\ S_n/m) = m(h - th_1 - \mu t\, h_2) + nh_1 + S_n h_2 + \cdots,$$

where h_i denotes differentiation with respect to the ith argument, and h, h_1, and h_2 are evaluated at $(t, \mu t)$. The heuristic reasoning following (2.27) suggests that the higher order terms play no role in determining the asymptotic distribution of R_m, which is thus obtained by applying (2.8) to $nh_1 + S_n h_2$. Although this conjecture is correct and allows one to obtain easily the result claimed in the statement of Theorem 3.11, there are two technicalities making a rigorous proof more difficult. (i) Unlike the situation in Theorem 2.18, the $\bar{P}_\xi^{(m_1)}$-process $S_{m_1} - S_n$, $n = m_1$, $m_1 - 1, \cdots$ is *not* a random walk, so the renewal theorem is not directly applicable. (ii) Even if it were a random walk, the technical conditions of Lai and Siegmund (1977) are not fulfilled, and no minor change in their argument produces an appropriate result.

Fortunately (ii) is solved by Hogan (1984), who considers nonlinear renewal theory for processes of the form (A.3). See Appendix 2. To circumvent (i), note that by (A.1) and (A.2) it suffices to evaluate

$$E_{\lambda,\xi}^{(m_1)}\{[(m - T^*)/T^*]^{1/2} \exp(-R_m^*);\ T^* \geq m_0\}$$
$$= E_{0,\lambda-\xi}^{(m_1)}\{[(m - m_1 + r)/(m_1 - r)]^{1/2} \exp(-R_m);\ r \leq m_1 - m_0\}$$

and then integrate out λ. For λ of the form

$$\lambda = m\xi_0/(1 - t_1) + \eta m_1^{1/2},$$

one can easily calculate the likelihood ratio of $z_1, \cdots, z_n$ under $P_{0,\lambda-\xi}^{(m_1)}$ relative to the unconditional probability P_μ ($\mu = \xi_0/(1 - t_1)$), which has essentially the same drift per unit

time, to obtain

$$E_{0,\lambda-\xi}^{(m_1)}\{[(m-m_1+r)/(m_1-r)]^{1/2}\exp(-R_m);\ r\leq m_1-m_0\}$$

$$= E_\mu\Bigg\{\{[m_1(m-m_1+r)]^{1/2}/(m_1-r)\}\exp\Bigg[-R_m-\frac{1}{2}(S_r-\mu r)^2/(m_1-r)$$

$$+ 2\eta(S_r-\mu r)/m^{1/2}-\frac{1}{2}\eta^2 r/(m_1-r)\Bigg];\ r\leq m_1-m_0\Bigg\}.$$

It is now straightforward, but tedious to use the asymptotic degeneracy of r/m, the asymptotic normality of $[S_r-\mu r]/r^{1/2}$, its asymptotic independence from R_m (as in Lemma 2.16) and the P_μ-limiting distribution of R_m given by Hogan's (1984) nonlinear renewal theorem to evaluate this expectation and hence complete the proof of Theorem 3.11.

Appendix 2
Nonlinear Renewal Theory

In Section 2.3 the renewal theorem was used to approximate the distribution of the excess of a random walk at its first passage across a linear boundary. In Theorems 2.18 and 3.11 similar problems arose with regard to first passages to nonlinear boundaries. In this appendix we survey the appropriate nonlinear renewal theory.

The problem is complicated by the fact that the stopping time and the excess over the boundary usually can be defined in more than one way. Conceptually the simplest situation involves a stopping time of the form

$$(A.4) \qquad T = T_m = \inf\{n : S_n \geq mc(n/m)\}.$$

Here $c(\cdot)$ is a positive continuous function, S_n, $n = 1, 2, \cdots$ is a random walk with positive drift $\mu = E(S_1)$, and we assume that the ray μt crosses the curve $c(t)$ at exactly one point, $\bar{t}$, near which $c(\cdot)$ is twice continuously differentiable. The stopping rules r^* and r introduced in the proofs of Theorems 2.18 and 3.11 respectively are both essentially of this form.

It follows from an argument based on the strong law of large numbers that $T_m/m \to \bar{t}$ with probability one as $m \to \infty$. Since as $m \to \infty$ the curve $mc(n/m)$ for n close to $m\bar{t}$ flattens out, it is natural to conjecture that the excess over the curved boundary, $R_m = S_T - mc(T/m)$, converges in law to the same limit as the excess over the tangent to $mc(\cdot)$ at the point $\bar{t}$, which is given by (2.8) with $\tilde{S}_n = S_n - nc'(\bar{t})$.

Although the conjecture of the preceding sentence is true under quite general conditions, in special cases it follows from a somewhat different result, which is considerably easier to prove. When possible, it is convenient to rewrite (A.4) in the form

$$(A.5) \qquad T = T_a = \inf\{n : ng(n^{-1}S_n) \geq a\}$$

for suitable g and a (depending on m). For example, for $c(t) = c_0 t^\gamma$ $(0 < \gamma < 1)$ in (A.4), we find that $g(x) = (x^+)^{(1-\gamma)^{-1}}$ and $a = c_0^{(1-\gamma)^{-1}} m^{1-\gamma}$. For the stopping time (A.5), the excess over the boundary is $\tilde{R}_a = Tg(S_T/T) - a$. A Taylor series expansion of g yields

$$ng(S_n/n) = ng(\mu) + (S_n - n\mu)g'(\mu) + (S_n - n\mu)^2 g''(\varsigma_n)/2n,$$

where ς_n satisfies $|\varsigma_n - \mu| \leq |n^{-1}S_n - \mu|$. If $g(\mu) > 0$ (as we shall assume), the linear part of $ng(S_n/n)$ is a random walk increasing at a rate proportional to n, whereas the quadratic part is essentially constant. This leads one to suspect that the limiting distribution of $\tilde{R}_a$ is given by (2.8) with $\tilde{S}_n = ng(\mu) + (S_n - n\mu)g'(\mu)$. This conjecture is also true and has been given an abstract formulation by Lai and Siegmund (1977). They consider stopping rules of the form

$$(A.6) \qquad \qquad \tilde{T} = \inf\{n : \tilde{S}_n + \tilde{\eta}_n \geq a\},$$

where $\tilde{S}_n$, $n = 1, 2, \cdots$ is a non-arithmetic random walk with positive mean $\tilde{\mu} = E\tilde{S}_1$ and $\tilde{\eta}_n$ changes sufficiently slowly in a sense made precise below that it plays no role in determining the limiting distribution of $\tilde{S}_{\tilde{T}} + \tilde{\eta}_{\tilde{T}} - a$ as $a \to \infty$. A typical application is to prove that the limiting distribution of $\tilde{R}_a$ is as indicated above. Lai and Siegmund also apply their result to approximate the significance level, (1.2) with $\mu = 0$, of a repeated significance test. Lalley (1983) extends the Lai-Siegmund method to the much more difficult case of multiparameter exponential families.

Although the stopping rule (1.4) of a repeated likelihood ratio test in an exponential family is of the form (A.5), the arguments used in this paper to prove Theorems 2.18 and 3.11 introduce auxiliary stopping rules which cannot be put into that form. Actually r^* defined in (2.28) to prove Theorem 2.18 is almost of the abstract form (A.6) with $\tilde{S}_n = \frac{1}{2}\mu_1^2 \, \xi_0^{-1} + S_n$, $\tilde{\eta}_n = \frac{1}{2}m^{-1} \, \xi_0^{-1} \, S_n^2$, and $a = \frac{1}{2}\xi_0^{-1} \, m(\mu_1^2 - \xi_0^2)$, except that Lai and Siegmund do not permit $\tilde{\eta}_n$ to depend on a. A suitable essentially trivial extension, modeled on Woodroofe's (1982) reformulation of the Lai-Siegmund result, is given below. This performs the dual function of completing the proof of Theorem 2.18 and of explaining the general nature of this class of results. Then we discuss briefly the method used by Hogan (1984) to deal with stopping times of the form (A.3) or (A.4).

Theorem A.7. Let $\tilde{T}$ be defined by (A.6), where $\tilde{S}_n$, $n = 1, 2, \cdots$ is a non-arithmetic random walk with positive mean $\tilde{\mu} = E(\tilde{S}_1)$ and for all n $\tilde{\eta}_n = \tilde{\eta}_n(a)$ is a measurable function of $\tilde{S}_1, \cdots, \tilde{S}_n$. Suppose also that for each $\lambda > 0$

$$a^{-1} \max_{1 \leq n \leq a\lambda} |\tilde{\eta}_n| \xrightarrow{P} 0 \qquad (a \to \infty)$$

$$\times = \tilde{S}_n + \tilde{\eta}_n$$

$$\circ = \tilde{S}_{n_1} + \tilde{\eta}_{n_1} + \tilde{S}_n - \tilde{S}_{n_1}$$

Figure 7

and for each λ, $\epsilon > 0$ there exists $\delta = \delta(\lambda, \epsilon)$ such that

$$(A.8) \qquad \max_{n \leq a\lambda} P\left\{ \max_{1 \leq k \leq a\delta} |\tilde{\eta}_{n+k} - \tilde{\eta}_n| > \epsilon \right\} < \epsilon.$$

Then as $a \to \infty$, for all $x \geq 0$

$$P\{T < \infty, \ \tilde{S}_T + \tilde{\eta}_T - a \leq x\} \to H(x),$$

where H is defined to be the right hand side of (2.8).

With the help of Theorem A.7 one can easily complete the proof of Theorem 2.18. The critical condition in the statement of Theorem A.7 is (A.8). The method of proof involves conditioning on $\tilde{S}_{n_1} + \tilde{\eta}_{n_1}$, where n_1 is chosen so that $\tilde{S}_{n_1} + \tilde{\eta}_{n_1}$ is already close enough to a that by (A.8) $\tilde{\eta}_n$ is constant (to within ϵ) for all $n_1 \leq n \leq \tilde{T}$, but it is far enough from a that the renewal theorem applies to the random walk $\tilde{S}_n - \tilde{S}_{n_1}$, $n = n_1 + 1, \cdots$. Hence except for an event of arbitrarily small probability $\tilde{S}_n + \tilde{\eta}_n$ and $\tilde{S}_{n_1} + \tilde{\eta}_{n_1} + (\tilde{S}_n - \tilde{S}_{n_1})$ cross a at the same time and have the same excess, to within ϵ. See Figure 7. The renewal theorem gives the indicated limiting distribution of excess over the boundary for the second process and hence for the process of interest.

It is easy to see where this argument runs into difficulty in dealing with a stopping rule of the form (A.4) or what is essentially the same, (A.3). If we assume that $\overline{E S_1^2} < \infty$, the variability in S_n is $O(n^{1/2})$. Hence to have probability close to one that $S_{n_1} < mc(n_1/m)$, one must choose $n_1 = m\bar{t} - Km^{1/2}$ for some *large* value of K. But a Taylor expansion shows that $mc(n/m)$ and the tangent $mc(\bar{t}) + c'(\bar{t})(n - m\bar{t})$ are essentially the same only for $|n - m\bar{t}| \leq \delta m^{1/2}$ for *small* δ. To circumvent this difficulty one can introduce the auxiliary stopping time

$$T_1 = \inf\{n : S_n \geq mc(n/m) - \delta m^{1/2}\}.$$

From the fact that $m^{-1}T_m \to \bar{t}$ and the assumption $E S_1^2 < \infty$, it follows that with probability approaching one

$$S_{T_1} - [mc(T_1/m) - \delta m^{1/2}] \leq \max_{1 \leq i \leq T_1} (S_i - S_{i-1})^+ + \sup |c'(t)|$$

is $o(m^{1/2})$ and hence $mc(T_1/m) - S_{T_1}$ is large. Moreover, during the approximately $\delta m^{1/2}/ [\mu - c'(\bar{t})]$ additional steps the random walk requires to cross the curve, the distance between the curve and its tangent is small, provided δ is small. Hence the Lai-Siegmund argument with the random time T_1 in place of n_1 shows that the time at which the random walk crosses the curve and the excess over the curve are with high probability equal to the time it crosses the tangent and almost equal to the excess over the tangent. Thus the nonlinear problem is reduced to a linear one having an answer given by (2.8) with $\tilde{S}_n = S_n - nc'(\bar{t})$.

This argument is easily made precise and also extended along the lines of Lemma 2.16. The result provides an appropriate tool for completing the proof of Theorem 3.11, or of Theorem 2.18 for that matter.

Hogan (1984) develops a much more sophisticated version of the argument for stopping times of the form (A.3). He does not require that $E S_1^2 < \infty$, and in his definition of T_1 he replaces $\delta m^{1/2}$ by a large constant K. This minimizes the smoothness conditions imposed on h (or $c(\cdot)$). More importantly, however, Hogan's method also proves a nonlinear renewal theorem in problems scaled for a diffusion approximation, where the methods described here and also Woodroofe's (1976a) method fail completely.

It would be interesting to give an abstract formulation of Hogan's result for a stopping

time as in (A.6), since the method seems much more general than the cases actually covered by Hogan's theorems.

References

Adler, R. and Brown, L. (1984). Tail behavior for suprema of empirical processes, unpublished manuscript.

Anderson, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size, *Ann. Math. Statist.* **31**, 165–197.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193–212.

Armitage, P. (1975). *Sequential Medical Trials*, 2nd ed., Oxford:

Berk. R. and Brown, L. (1978). Sequential Bahadur efficiency, *Ann. Statist.* **6**, 567–581.

BHAT (β-blocker heart attack trial research group). (1981). β-blocker heart attack trial – design features, *Controlled Clin. Trials*, **2**, 275–285.

BHAT (β-blocker heart attack trial research group). (1982). A randomized trial of propranolol in patients with acute myocardial infarction, *J. Amer. Med. Assoc.* **147**, 1707–1714.

Bhattacharya, P. K. and Brockwell, P. J. (1976). The minimum of an additive process with applications to signal estimation and storage theory, *Z. Wahrschein. verw. Geb.* **37**, 51–75.

Bickel, P. and Rosenblatt, M. (1973). Two-dimensional random fields, in *Multivariate Analysis - III*, P. K. Krishnaiah, ed., Academic Press, New York, 3–15.

Birnbaum, Z. W. and Tingey, F. H. (1951). One-sided confidence contours for probability distribution functions, *Ann. Math. Statist.* **22**, 592–596.

Borovkov, A. A. (1962). New limit theorems in boundary problems for sums of independent terms, *Selected Translations in Mathematical Statistics and Probability* **5**, 315–372.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time, *J. Roy. Statist. Soc. B* **37**, 149–192.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time, *Ann. Math. Statist.* **35**, 999–1028.

Cobb, G. W. (1978). The problem of the Nile: conditional solution to a change point problem, *Biometrika* **62**, 243–251.

Cox, D. R. (1972). Regression models and life tables (with discussion), *J. Roy. Statist. Soc. B* **34**, 187–220.

Daniels, H. (1974). The maximum size of a closed epidemic, *Adv. Appl. Prob.* **6**, 607–621.

DeLong, D. M. (1981). Crossing probabilities for a square root boundary by a Bessel process, *Commun. Statist. – Theor. Math.*, **A10**, 2197–2213.

DeMets, D., Hardy, R., Friedman, L., and Lan, K. K. G. (1984). Statistical aspects of early termination in the beta-blocker attack trial, *Controlled Clinical Trials*, **4**.

Durbin, J. (1981). The first pasage density of a Gaussian process to a curved boundary, unpublished manucript.

Feller, W. (1972). *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley and Sons, New York.

Gardner, L. A., Jr. (1969). On detecting changes in the mean of normal variates, *Ann. Math. Statist.* **40**, 116–126.

Gill, R. D. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre-Tracts **124**, Mathematisch Centrum, Amsterdam.

Halpern, J. (1982). Points of interest on the grand tour, Stanford University Technical Report.

Haybittle, J. L. (1971). Repeated assessments of results in clinical trials of cancer treatment, *British J. of Radiology*, **44**, 793–797.

Hinkley, D. V. (1970). Inference about the change point in a sequence of random variables, *Biometrika* **57**, 1–17.

Hinkley, D. V. (1972). Time ordered classification, *Biometrika* **59**, 509–523.

Hogan, M. (1984). Problems in boundary crossings for random walks, Stanford University dissertation.

Hu, Inchi (1985). On repeated significance tests. Stanford University dissertation.

Jennen, C. and Lerche, R. (1981). First exit densities of Brownian motion through one-sided moving boundaries, *Z. Wahrsch. verw. Gebiete*, **55**, 133–148.

Kendall, D. G. and Kendall, W. S. (1980). Alignments in two-dimensional random sets of points, *Adv. Appl. Prob.* **12**, 380–424.

Lai, T. L. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis I, *Ann. Statist.*, **5**, 946–954.

Lalley, S. (1983). Repeated likelihood ratio tests for curved exponential families, *Z. Wahrsch. verw. Gebiete*, **62**, 293–321.

Levin, B. and Kline, J. (1984). The cusum test of homogeneity, with an application to spontaneous abortion epidemiology, Columbia University, preprint.

Lorden, G. (1971). Procedures for reacting to a change in distribution, *Ann. Math. Statist.*, **42**, 1897–1908.

Miller, R. G. (1970). Sequential rank tests – one sample case, *Proc. Sixth Berk. Symp. Math. Statist. and Prob.*, **1**, University of California Press, Berkeley, 97–108.

Miller, R. G., Jr. and Siegmund, D. (1982). Maximally selected chi square statistic, *Biometrics*, **38**, 1011-1016.

O'Brien, P. C. and Fleming, T.R. (1979). A multiple testing procedure for clinical trials, *Biometrics*, **35**, 549–556.

Page, E. S. (1954). Continuous inspection schemes, *Biometrika*, **41**, 100–115.

Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J., Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient, *Br. J. Cancer*, **34**, 585–612.

Pettit, A. N. (1980). A simple cumulative sum type statistic for the change-point problem with zero-one observations, *Biometrika* **67**, 79–84.

Pickands, J. (1969). Upcrossing probabilities for stationary Gaussian processes, *Trans. Amer. Math. Soc.* **145**, 51–73.

Pocock, S. (1977). Group sequential methods in the design of an analysis of clinical trials, *Biometrika*, **64**, 191–200.

Pollak, M. (1985). Optimal detection of a change in distribution, to appear in *Ann. Statist.*

Qualls, C. and Watanabe, H. (1973). Asymptotic properties of Gaussian random fields, *Trans. Amer. Math. Soc.* **177**, 155–171.

Samuel-Cahn, E. (1974). Repeated significance tests II, for hypotheses about the normal distribution, *Commun. Statist.*, **3**, 711–733.

Samuel-Cahn, E. and Wax, Y. (1984). A sequential "gambler's-ruin" test for $p_1 = p_2$, when p_i are small, with a medical application, and incorporation of data accumulated after stopping, Hebrew University Technical Report.

Sellke, T. (1985). Evolution of the partial likelihood over time, Purdue University Technical Report.

Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model, *Biometrika*, **70**, 315–326.

Sen, A. and Srivastava, M. S. (1975). On tests for detecting changes in means, *Ann. Statist.* **3**, 98–108.

Shiryayev, A. N. (1963). On optimal methods in earliest detection problems, *Theory of Probab. and Its Appl.*, **8**, 26–51.

Siegmund, D. (1975). Importance sampling in the Monte Carlo study of sequential tests, *Ann. Statist.*, **4**, 673–684.

Siegmund, D. (1977). Repeated significance tests for a normal mean, *Biometrika*, **64**, 177–189.

Siegmund, D. (1978). Estimation following sequential tests, *Biometrika*, **65**, 341–349.

Siegmund,D. (1982). Large deviations for boundary crossing probabilities, *Ann. Probab.*, **10**, 581–588.

Siegmund, D. (1984). Corrected diffusion approximations and their applications, *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, L. LeCam and R. Olshen, eds., Wadsworth, Belmont.

Siegmund, D. and Yuh, Yih-Shyh (1982). Brownian approximations to first passage probabilities, *Z. Wahrsch. verw. Gebiete*, **59**, 239–248.

Stam, A. J. (1968). Two theorems in r-dimensional theory, *Z. Wahrsch. verw. Gebiete*, **10**, 81–86.

Woodroofe, M. (1976a). A renewal theorem for curved boundaries and moments of first passage times, *Ann. Probab.*, **4**, 67–80.

Woodroofe, M. (1976b). Frequentist properties of Bayesian sequential tests, *Biometrika* **63**, 101–110.

Woodroofe, M. (1978). Large deviations of the likelihood ratio statistic with applications to sequential testing, *Ann. Statist.*, **6**, 72–84.

Woodroofe, M. (1979). Repeated likelihood ratio tests, *Biometrika*, **66**, 453–463.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Society for Industrial and Applied Mathematics, Philadelphia.

Woodroofe, M. and Takahashi, H. (1982). Asymptotic expansions for the error probabilities of some repeated significance tests, *Ann. Statist.*, **10**, 895–908.

Worsley, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability, *Biometrika* **70**, 455–464.

Yuh, Yih-Shyh (1982). Second order corrections for Brownian motion approximations to first passage probabilities, *Adv. Appl. Probab.*, **14**, 566–581.

END

FILMED

4-85

DTIC